ULTRASOUND AND ACOUSTIC ANALYSIS OF LINGUAL MOVEMENT IN TEENAGERS WITH CHILDHOOD APRAXIA OF SPEECH, CONTROL ADULTS AND TYPICALLY DEVELOPING CHILDREN

TANJA KOCJANČIČ

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

QUEEN MARGARET UNIVERSITY

2010

Abstract

Childhood apraxia of speech (CAS) is a neurological motor speech disorder affecting spatiotemporal planning of speech movements. Speech characteristics of CAS are still not well defined and the main aim of this thesis was to reveal them by analysing acoustic and articulatory data obtained by ultrasound imaging. Ultrasound recording provided temporal and articulatory measurement of duration of syllables and segments, amount and rate of tongue movement over the syllables and observation of the patterns of tongue movement. Data was provided by three teenagers with CAS and two control groups, one of ten typically developing children and the other of ten adults. Results showed that, as a group, speakers with CAS differed from the adults but not from the typically developing children in syllable duration and in rate of tongue movement. They did not differ from either of the control groups in amount of tongue movement. Individually, speakers with CAS showed similar or even greater consistency on these features than the control speakers but displayed different abilities to adapt them to changes in the syllable structure. While all three adapted syllable duration and rate of tongue movement in the adult-like way, only two showed mature adaptation of segment durations and of the amount of tongue movement. Observing patterns of tongue movement showed that speakers with CAS produce different patterns than speakers in the control groups but are at the same time, like adults, very stable in their articulations. Also, speakers with CAS may move their tongues less in the oral space than speakers in the control groups. The differences between the control groups were similar to those found in previous studies. The results provide support for the validity of the methods used, new information about CAS and a promising direction for future research in differential diagnostic and therapy procedures.

Acknowledgements

First of all, I would like to thank my director of studies, Dr Nigel Hewlett, whose guidance and support from the initial to the final level enabled me to progress in my research and to develop an understanding of the subject. This thesis would additionally not be possible without the contribution of my other supervisors: Dr Marianne Pouplier, Dr Robin Lickley and Prof Fiona Gibbon. They have all taken time to answer my questions, read the reports at various levels of the research and contribute helpful suggestions.

This thesis would not have been possible without the participants and their families. I am grateful for their time and willingness to take part. I would also like to thank Prof Anne O'Hare and Mrs Susan Davidson for invaluable help with the recruitment of speakers with childhood apraxia of speech and with the application for the NHS ethical approval.

I am indebted to the members of Speech and Hearing Sciences at Queen Margaret University for support and help at any time I was stuck. Especially big thank you to Steven Cowen for helping me with the technical details and troubleshooting during the recordings, to Dr Alan Wrench for all the advice about ultrasound recording and prompt fixing of the software bugs and to the members of the Articulate Assistant Advance group for interesting discussions.

Great thank you also to the members of the Centre for Speech Technology Research at the University of Edinburgh who made me very welcomed and enabled me to learn some very different things about speech.

I owe my deepest gratitude to the colleagues from both departments who formed the Edinburgh Speech Science and Research group and gave me the opportunity to join them and start doing research in the area that interests me the most. Additionally, I am grateful to the Marie Curie European Commission network for providing the necessary funding.

I would further like to thank all of my office mates for making day-to-day coping with the PhD easier and much more fun.

I am very grateful to my family: mami, oči, mama, babi in teti. They always understood how much it means to me to continue with further education and were always supporting my decision.

Finally, I am also thanking Jano, not only for all the help with Matlab scripts, but for making my PhD-study time even nicer and memorable. It was fun to be on the PhD journey together and reaching the goal at the same time.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Childhood apraxia of speech (CAS) was until recently one of the most controversial speech disorders. Speakers with CAS have always been reported to present a number of speech characteristics: segmental errors, impaired prosody, increasing difficulties with increased length of speech material, groping movements, inconsistency of productions on different attempts, better speech perception than production, and no problems executing non-speech oral movements. The problem was, however, that none of these characteristics are unique to CAS and that not all speakers diagnosed with the impairment exhibited all of them. The understanding of CAS was further affected by lack of agreement about the underlying cause of such speech profile. It was suggested that CAS may result from a pure motor deficit, pure phonological deficit or a combination of the two. Additionally, nothing was known about the cause of the impairment and to complicate its understanding even more, different names were used to describe what was believed to be the same impairment. All these uncertainties made diagnosis of CAS rather difficult, causing misdiagnoses, and consequently inappropriate therapy procedures, raising doubt about the reported cases of CAS truly having the impairment, and making any firm conclusions about CAS very difficult, if not impossible. The definition of CAS was set only very recently in a technical report published by the American Speech-Language-Hearing Association (ASHA) (American Speech-Language-Hearing Association, 2007). They defined the impairment as CAS, confirmed its neuro-motor origin and main problems in spatiotemporal planning of speech sequences. However, there was no conclusion about the typical speech characteristics or the cause of CAS.

Because of the relativity poor understanding of CAS, research has been focused mainly on discovering its main speech characteristics, and a subset of those that would make differential diagnostic easier. As in the case of most speech impairments, the first attempts were focused on analysing the number and type of segmental errors, followed by the acoustic profiles, including stress patterns and temporal properties. The few studies that focused on the characteristics of articulatory movements were based mainly on acoustic analysis and only two on articulatory data. However, because CAS is a motor speech impairment, it seems reasonable to research it more extensively with articulatory methods. They can potentially reveal not only some new information about speech in CAS, but, even more importantly, some unique characteristics that are not present in phonological disorders with similar output features. This would allow earlier and more reliable diagnosis, application of correct therapy procedures and, eventually, a better and faster correction.

The main aim of this thesis was to investigate speech characteristics of CAS, particularly tongue movements and temporal features, by using ultrasound imaging. If speakers with CAS have impaired spatiotemporal planning of speech movements, this could be observed in different temporal features, especially when adapting durations in different syllable structures, and in deviant characteristics of tongue movements. Although CAS is a motor speech disorder it has not been extensively researched by applying articulatory methods. Because the tongue is one of the busiest articulators in speech, moving in a number of directions and adopting different shapes, and because it is involved in the production of most speech sounds, it seems an informative object of research. The reasons that tongue movements in CAS, or other speech disorders, have not been widely researched probably lie in the fact that the tongue is positioned deep in the oral cavity and for that reason it is not visible during speech, and that all articulatory methods are more or less invasive for the participants, which might not be suitable for speakers with impaired speech, particularly at a young age. The solution to this problem might be the application of ultrasound imaging. Ultrasound is safe and, of all the articulatory methods, the least invasive method to observe and measure tongue movements. In its most basic application, the ultrasound probe simply has to be placed under the speaker's chin and already the tongue surface can be observed. Additionally, recent developments in ultrasound imaging of the tongue allow synchronisation of the ultrasound and audio signals, enabling simultaneous acoustic and articulatory analysis of the same data. Because of its relative ease of usage, ultrasound seems to be suitable for research of CAS.

Research presented in this thesis is thus a first attempt at quantitative and qualitative description of tongue movements in CAS from ultrasound images and an attempt to combine acoustic and articulatory data. The production of speakers with CAS are compared to two control groups: adults with normal speech (referred to only as adults throughout this thesis) and typically developing children. In this way, their characteristics can be compared to both mature and immature, but still developing, speech production systems. Additionally, the results of the control groups can be compared to a number of earlier studies of temporal and articulatory characteristics of adults and children in order to validate the methodological procedures used in the study presented here.

Chapter 2 is a description of background to this study. First, it presents how our understanding of CAS has changed over time by outlining the main speech characteristics and stressing the complexity of the impairment. After identifying that temporal and articulatory characteristics of speech could present some unique CAS features, the next sections of this chapter address both of this characteristic in adults and typically developing children. The last section of the chapter gives an overview of ultrasound imaging, its application to the field of speech science and the ultrasound system used at Queen Margaret University.

Chapter 3 is a description of the methodology, giving details about participating speakers, presenting speech material, and explaining recording procedures and data analysis.

Results can be seen in Chapter 4. Sections 4.1, 4.2 and 4.3 show group results of syllable and segment durations, of amount of tongue movement over syllables, and of combined articulatory and acoustic data, respectively. The next section, 4.4, presents tongue movement patterns of the three participating speakers with CAS and three speakers from each of the control groups. Because of small number of speakers with CAS, results of individual speakers with CAS on all of the measures are given in section 4.5.

Finally, Chapter 5 is a general discussion of the results obtained. First, it provides a critical evaluation of the statistical methods, followed by conclusions about the speech characteristics of CAS as compared to the two control groups. The chapter additionally provides discussion about the shortcomings of the study and gives suggestions for future work.

Chapter 2

Background

2.1 Childhood apraxia of speech

The first description of what is now known as CAS was published in 1954. Morley et al. (1954) described speakers who were able to produce non-speech oral movements but had difficulties realising speech tasks and named the disorder dyspraxic dysarthria. Since then CAS has turned into quite a controversial speech impairment with researchers and clinicians disagreeing about the name, etiology, underlying deficit, characteristics and even about its existence. The proposed explanations ranged from CAS being just a more severe form of speech delay of unknown origin, a special kind of phonological disorder, a pure motor disorder, a children's version of acquired apraxia of speech observed in adults or a combination of phonological and motor speech impairments. In 1981 Guyette and Diedrich even called it "a label in search of a population" (p.39) suggesting that the name of an impairment has been created without speakers to represent it. The story of CAS was further complicated by different names used to describe the same disorder. The impairment was called developmental dyspraxia, developmental verbal dyspraxia/apraxia, articulatory apraxia, developmental apraxia of speech, childhood apraxia of speech, and suspected childhood apraxia of speech (sCAS).

The situation was somewhat resolved in 2007 when ASHA recommended classifying the disorder as childhood apraxia of speech and defined it as a

> "neurological childhood (paediatric) speech sound disorder in which the precision and consistency of movements underlying speech are impaired in the absence of neuromuscular deficits (e.g., abnormal reflexes, abnormal tone). CAS may occur as a result of known neurological impairment, in association with complex neurobehavioral disorders of known or unknown

> origin, or as an idiopathic neurogenic speech sound disorder. The core impairment in planning and/or programming spatiotemporal parameters of movement sequences results in errors in speech sound production and prosody." (American Speech Language Hearing Association, p.3-4).

The ASHA description is based on the research on CAS performed between 1995 and 2007 which was focused on revealing either the main characteristics of the disorder, or diagnostic markers used to differentiate between CAS and other disorders of speech production with unknown origin. It provides some information about the nature of the disorder, but it does not inform on the exact origin of CAS or its characteristics. The description is thus by no means exhaustive and final. However, it provides information that can serve as a basis of future research and clinical practice.

Before addressing CAS in any more detail, one of the main issues affecting any conclusions about the disorder has to be pointed out: how to be sure that speakers with CAS who participated in any of the CAS studies truly have CAS and not any other speech disorders? The selection process is made difficult because the list of CAS characteristics is not set, because not all children with CAS exhibit all of the characteristics, and because proposed CAS characteristics are observed in other speech sound disorders as well (McCabe et al., 1998). The most common way of selecting participants is referral from speech and language therapists. This approach has put in some doubt by Forrest (2003) who showed that speech therapists use a very wide range of characteristics to diagnose CAS. Forrest asked 75 speech and language therapists who all had at least some experience with CAS to name three characteristics that are necessary for a diagnosis of CAS. The therapists named 50 different characteristics out of which six occurred in 51% of all responses, and 20 characteristics that were named only once. The top six necessary characteristics were: inconsistent productions (14.1%), general oral-motor difficulties (9.3%), groping (7.9%), inability to imitate sounds (7.5%), increased errors with increased utterance length (6.6%) and poor sequencing of sounds (6.2%). Similar observations were made by Davis et al. (1998) who tested 22 children referred to their group as having CAS but confirmed the diagnosis only in four.

Despite such a big variety of diagnostic characteristics across the therapists, it would be expected that speech and language therapists who have years of experience diagnosing and treating not just CAS but different speech disorders are more reliable in recognising CAS and not making a false diagnosis.

Some of the researchers investigating CAS have tested referred speakers on a number of tests to confirm the diagnosis. They mostly followed the whole set or subset of characteristics proposed by Davis et al. (1998) to recognise CAS and distinguish it from other speech and language disorders. These characteristics are: limited consonant and vowel repertoire, frequent omission errors, frequent vowel errors, inconsistent errors, altered suprasegmental characteristics, number of errors increasing with increased length of spoken output, difficulty imitating words and phrases with groping gestures, predominant use of simple syllable structures, impaired volitional oral movements, reduced expressive compared to receptive language skills, and reduced diadochokinetic rates.

Following these two selection procedures enabled better reliance that the children participating in studies of CAS truly have the disorder and that their speech characteristics present characteristics of CAS. This allowed making conclusions about CAS, expand our knowledge of it, and finally, led to agreement about its motor deficit origin. The following sections describe the earlier research of CAS and provide more information about the disorder, with an overview of the origins, prevalence and gender structure, and main speech characteristics.

2.1.1 Origins of CAS

The lack of clear origin of CAS has always been a problem in describing and diagnosing the disorder and is evident in the ASHA definition as well. Only a very small number of studies have addressed the question so far and they have failed to find definite support for either a genetic or neurological cause. Most evidence about the origin of CAS comes from a number of studies of the British KE family (Fisher et al., 1998; Lei et al., 2001; Vargha-Khadem et al., 2005). Four generations of the family showed a high occurrence of language and speech disorders, including CAS-like characteristics. Genetic studies have revealed that 15 members of the family affected by CAS had also a deficit in the FOXP2 gene (American Speech-Language-Hearing Association, 2007). The same gene was also found to be affected in a mother and a daughter with CAS studied by Shriberg et al. (2006). However, no clear answer about the genetic origins of CAS was found in a study by Lewis et al. (2004). They evaluated family members of 22 children with CAS, and found only two siblings who shared the same speech disorders. However, they did find that 86% of children had at least one nuclear family member affected by a speech and/or language disorder, and 59% had at least one affected parent. Such results did not allow the authors to make claims about the family aggregation of CAS but it did show evidence for a genetic origin of speech and/or language disorders since family members of children with CAS were more often affected than family members of children with other speech, and speech and language disorders.

Data on neurological profiles of CAS is even more sparse. Although members of the KE family have been widely investigated and have shown several neurological differences in morphology and functioning, the findings are difficult to relate to CAS only since the speakers had a greater extent of speech problems than typically observed in children with CAS. Additionally, some evidence was also given that CAS is more likely to occur in complex neurobehavioral disorders such as autism, epilepsy, fragile X syndrome, galactosemia, Rett syndrome and chromosome translocations (American Speech-Language-Hearing Association, 2007).

2.1.2 Prevalence and gender structure

Because of the problems diagnosing CAS, there is no certain information about the prevalence of the impairment and about the gender structure.

The first estimation of prevalence suggested that around 1% of all children are affected by CAS (1.3% by Morley 1972, and 1% by Yoss 1975, both cited in Shriberg et al. 1997, p.277). Shriberg and Kwiatkowski (1994, cited in Shriberg et al. 1997, p.277) proposed a much lower estimate of 0.125%, while Delaney and Kent (2004, cited in American Speech Language Hearing Association, p.5) reported that 3.4%-4.3% of all children are affected. However, it is not clear whether such a high estimate is truly the result of an increase of CAS cases among children, a result of more frequent evaluations of speech development due to demands for early diagnostic, or the result of more wrongly diagnosed children (American Speech-Language-Hearing Association, 2007).

The gender structure of children with CAS has been evaluated mainly from the reported studies and not from a specific reviews of treated clients in a clinic. After reviewing published studies, Halle et al. (1993, cited in Shriberg et al. 1997, p.276) reported that the male:female ratio is at 3:1. A review of 30 studies (published from 1993 on) presented in this thesis showed that 293 children with CAS or suspected CAS (sCAS) took part. 25 of these studies also give information about the gender of participants. The percentage of males in these individual studies ranged from 33% to 100%. When taken all together, out of 240 children, 176 (73%) were males, and 64 (27%) females. This supports the notion that about three times more boys are diagnosed than girls. However, I would like to stress that some authors published several papers from which it is not always possible to ascertain whether participants were always the same speakers or different ones.

2.1.3 Speech characteristics of CAS

As previously mentioned, research in CAS has been, and still is, orientated mainly at uncovering diagnostic markers that are typical for CAS and will allow making a quick and correct diagnosis which, in turn, is more likely to result in better correction of speech features, improved communication skills and lesser influence on other levels of child development. The studies published so far can be divided into those investigating speech sound errors, stress, temporal properties, articulatory movements and those seeking a relation between speech perception and production. Reviewing earlier studies is important for better understanding of CAS. On one hand, they show how our understanding of CAS has changed throughout the years, but on the other hand also present the complexity of the disorder and reveal the main reasons for existing problems in defining and characterising it. The following five sections give a historical overview of uncovering the main speech characteristics of CAS.

2.1.3.1 Speech sound errors

Analysing speech sound errors is one of the first steps in describing speech disorders. They are usually very notable and relatively easy to evaluate, by describing the types of errors or counting correct and wrong realisations.

Vowel errors are often reported as one of the main characteristics of CAS. They are very notable in speech, particularly because they are rather unusual. Davis et al. (2005) reported that typically developing children are able to produce appropriate vowels by the age of 24 months, and that vowel errors are unlikely to be observed in older children.

The first study analysing vowel performance in CAS was done by Pollock and Hall (1991) who followed previously reported observations of vowel errors in children with CAS. They recruited five children with CAS, aged between 8;2 (years;months) to 10;9 years, and investigated their vowels in elicited single word productions. Vowels were divided into four subcategories: non-rhotic vs. rhotic and monophthongs vs. diphthongs. Analyses revealed that all children had problems with rhotic vowels and diphthongs, and that accuracy for non-rhotic vowels ranged between 56% and 96%

for individual speakers. In spite of a high between-speaker variability in the amount of errors, children with CAS still showed similar error patterns. The most common errors included the substitutions of laxing, backing, lowering and tensing, diphthong reduction, vowel harmony and derhotisation.

Similar results were obtained in a longitudinal study investigating connected speech of three sCAS speakers assessed at three yearly intervals, from the age of 4;6, 5;10 and 5;6 years on (Davis et al., 2005). At each of the three assessment times children were able to produce all Standard American English non-rhotic vowel types. Production of rhotic vowels was observed only in a few instances. These results showed that CAS speakers are able to produce almost all vowels and thus do not have a simple delay in vowel development. However, they did show a high number of vowel errors, resulting in much lower accuracy levels than what was expected for their age. The lowest accuracy for all three children was observed for the rhotic vowels. Over time, two of the speakers showed an increase in the accuracy for all vowels measured, diphthongs, and non-rhotic vowels, while one of the children did not show any improvement. The latter participant was also the only one to show an increase in the number of vowel errors over time. Different changes over time do not necessarily reflect poorer abilities or more severe CAS of one child compared to the other two, but may result from inconsistency in speech production. At different times, the three children could perform differently. More similarities between speakers were observed in the type of errors. For all of them, substitution and derhotisation were the most frequent types of errors, suggesting that children with CAS are aware of the vowels and do not simply omit them but only simplify them to make the production easier. Recorded material also revealed that the children used predominantly short utterances and simple syllable structure. However, when testing for an effect of utterance length and syllable structure on vowel performance, no differences were found between different subgroups.

Differences between speakers were observed also in a study of deviant vowel productions of three German sCAS children, aged 5;9 - 6;3 years, and 21 matched controls (Blech et al., 2007). Significant differences in the formant values (F1, F2 and F3) were observed between two of the sCAS speakers and controls across all combinations of investigated categories: monophthongs and diphthongs, real words and non-words, single syllables and multisyllabic words. One of the sCAS children showed differences only for monophthongs in words, diphthongs in non-words, and both types of vowels in multisyllabic real and non-words. Taken together, these findings suggest that speakers with sCAS do not benefit from familiarity with words, vowel structure or word length in the same way as typically developing children do.

Consonant errors are usually a less reliable distinguishing feature of CAS because they are observed also in other speech impairments and in typical speech development. Typically developing children, for example, do not acquire all consonants and consonantal clusters until eight years of age (Templin, 1957; Smit et al., 1990). As shown in several studies investigating consonantal errors in CAS, children with CAS make more errors, especially in consonantal clusters, than typically developing children.

Thoonen et al. (1997) analysed errors in the production of multisyllabic real and non-words of 11 children with CAS (6;2 - 7;9 years) and 11 typically developing children (6;0 - 7;11 years). The most frequent error types in both groups were substitution, followed by distortions and omission. Both groups also had higher error rates on clusters than on singletons, with cluster reduction as the most frequent cluster error. Additionally, they both showed the same effect of segment position on the type of error. Syllable initial position was most commonly affected by substitutions, while omissions and cluster reductions were the most frequent error types in syllable final position. In contrast to typical speakers, CAS group showed more substitutions, omissions and distortions of singletons, more cluster reductions and more disfluencies. They were also differently affected by familiarity with the speech material. Control speakers made 12 times more errors on non-words than on real words, while CAS speakers' error ratio was only 1:2.2. Such a result suggests that familiarity had no effect on the error rates of CAS speakers and that speech characteristics of CAS can be tested on either familiar or non-familiar speech material.

Extensive problems with consonant cluster production was additionally demonstrated by Nijland et al. (2003b). Six Dutch children with CAS and six controls aged 4;11 - 6;10 took part in this study. They all uttered sequences with a syllable boundary before the cluster or with a syllable boundary separating the two consonants (/z/V1#[1]/sx/V2/t/ vs. /z/V1/s/#/x/V2/t/; V1 was /ə/ or /ɔ/, and V2 was /a/, /i/, or /o/). Analysis of syllable errors showed that four out of six children with CAS were more accurate at producing a /s/#/x/ sequence than the cluster. Further difficulties with cluster production were observed in three of the CAS children who were inserting pauses between the clustered consonants in some of the utterances. Such behaviour was not observed in the control group.

Another study of consonant errors in CAS (Marquardt et al., 2004) not only supported previous observations but also showed that problems with consonants are a

[1]# represents syllable boundary

relatively persistent feature of CAS. Three children with CAS were first recorded aged 4;6, 5;10 and 5;6 years, and than at about one year and two years later. After identifying the same target words at the three data collection time points, the authors measured token accuracy (percentage of tokens matching the target), target stability (target always realised in the same way vs. target realised in more than one way), total token variability (percentage reflecting number of variants over number of tokens) and error token variability (percentage reflecting number of error variants over number of tokens not matching the target). The three speakers showed different changes in the measurements over time, suggesting a high session-to-session variability. Overall, the children had higher token accuracy and target stability, and lower total variability at the last recording than at the first, but the change was not linear for all of them. The greatest difference was observed for error variability which increased for the first child, decreased for the second and stayed the same for the third. The authors point out that throughout the investigated period children were undertaking speech therapy which may have affected the results, especially when directed to improve token accuracy as it was for the first speaker. They also suggest that high variability on a performance task might be one of the key features of CAS and that multiple assessment is necessary to achieve the correct diagnosis.

When investigating speech disorders, it is not only necessary to describe them by comparing them to typical speech, but to other speech disorders as well. Knowing the difference between similar disorders is of great importance in the diagnostic phase. CAS can be often misdiagnosed as speech delay or phonological disorder (PD). For that reason, it is important to investigate the same speech characteristics in all these groups of speaker, in order to uncover those features that differentiate between them. Such an approach was used by Shriberg and colleagues (Shriberg and Aram, 1997a,b) who performed three studies with the aim of revealing one or more diagnostic markers for CAS. The first study involved 14 children with sCAS, aged 4;10 to 14;11, and 73 children with delayed speech, aged 3 - 13 years. Two more studies were then performed to validate the findings of the first one. One of them included 20 children with CAS (3;2 to 9;3), and 28 children with typical development and speech delay (3 - 19 years), and the second included 19 children with sCAS, aged 4;7 to 14;4. Analysis of the conversational data from the first study revealed that there were no differences between the sCAS and the speech delay groups on either frequency of errors, type of errors or error consistency. Error analysis alone is thus not enough to differentiate CAS from speech delay.

Similar conclusions were made in a study exploring the potential of error consistency as a diagnostic marker for CAS. Betz (2000, cited in Betz and Stoel-Gammon (2005)) compared word repetitions in five speakers with CAS and five speakers with PD. The main finding of this study was that although speakers with CAS made more errors than children with PD, the consistency of errors was the same across groups. Additionally, it was observed that the four children with the least consistent errors had also the most severe speech disorder as judged by intelligibility scores, disregarding the diagnoses of either CAS (three speakers) or PD (one speaker). The author concluded that error consistency is not directly related to CAS but it might be related to the severity of speech disorder.

Analyses of speech sound errors revealed that children with CAS make errors on vowels and consonants, particularly on more complex targets such as rhotic vowels, diphthongs and consonantal clusters. They make significantly more errors than typically developing children, but do not differ on all error measures from children with PD or speech delay. Furthermore, they showed high session-to-session and between-speaker variability. Children with CAS, additionally, did not benefit from familiarity with the words, and showed no effect of word length and syllable structure on vowel productions. They did, however, show an effect of syllable structure on consonants, as they performed worse on consonants in clusters than on singletons.

2.1.3.2 Stress

Another commonly reported speech characteristic in CAS is inappropriate, even stress.

The first studies addressing stress in CAS were the previously mentioned studies by Shriberg and colleagues (Shriberg and Aram, 1997a,b). They revealed that the only speech characteristic differentiating CAS from speech delay and typical development was inappropriate stress. Children with sCAS had significantly lower average scores of appropriate stress. Out of 13 children with sCAS included in the first of these studies, six had inappropriate stress, three questionable stress and only four appropriate stress. The two following evaluative studies confirmed these findings and implied that deficit in stress could be used as a diagnostic criteria to differentiate between sCAS and speech delay. Taken together, the three studies revealed that 52% of children with sCAS were recognised as having inappropriate stress but only 10% of children with speech delay. A question was immediately posed about why the remaining (48%) of children in the sCAS group showed appropriate stress and the authors suggested the existence of a CAS subgroup which is marked by stress deficit. However, this conclusion was later revised in another study by Shriberg et al. (2003a).

This time the authors tested the validity of a lexical stress ratio as a diagnostic marker for CAS. They recruited 11 children with sCAS, aged 3;3 - 10;10, and 24 children with speech delay, aged 3;4 - 12;0 years. All children produced eight familiar bisyllabic words in three stress patterns (trochaic (strong-weak), iambic (weak-strong) and spondee (strong-strong)). Based on the segmental measures of frequency area, amplitude area and duration of the eight trochaic words, the authors devised a formula to calculate lexical stress ratio. Ranking all the participants by their score revealed that out of six most extreme values (three highest and three lowest), five were from speakers with sCAS which is significantly higher than a chance level distribution would suggest. Such results imply that lexical stress ratio could be used to differentiate CAS from speech delay of unknown origin. Shriberg et al. (2003a) additionally proposed that due to the distribution of the scores of CAS speakers at either end of the ranking list, and due to the lack of clear discontinuity in the lexical stress ratio scores it is more likely that CAS has a motor origin than a phonological one. The results also showed that out of 11 participants with sCAS, perceptual analysis recognised 9 as having a stress deficit, but only five of them had atypical lexical stress ratio scores. These children were, however, not viewed as a subgroup of CAS. They were rather seen as displaying high day-to-day variability commonly observed in children with CAS.

A similar discrepancy between measured and perceived stress as in the latter study was also observed by Munson et al. (2003). They tested the potential of stress as a diagnostic marker to differentiate between CAS and PD on ten children aged 3;9 to 8;10 years, five in each of the two experimental groups. Children first produced non-words with trochaic and iambic stress pattern. Realisations were perceptually evaluated for accuracy of stress pattern. Additionally, acoustic measures of vowel duration, fundamental frequency at vowel midpoint, timing of the fundamental frequency at the vowel onset, and intensity at vowel midpoint were obtained. Comparing acoustic measures between groups revealed an interesting result. No group differences on any of the measures were observed as both groups were able to produce acoustic differences between stressed and non-stressed syllables. However, perceptual evaluation still resulted in sCAS children being less frequently judged to use a correct stress pattern. Post-hoc regression analysis showed that listeners' judgements were influenced by perceived duration and fundamental frequency only, although there were no significant differences in these two parameters between the groups.

Following research on stress impairment in CAS, it seems that stress is a better diagnostic marker than an error profile. Ability to use appropriate stress does not only differentiate children with CAS from typically developing children, but also from children with speech delay. However, just in the case of speech errors, high between speaker and session-to-session variability was observed, suggesting that children suspected of having CAS have to be evaluated more than once. Despite the good prospect for using stress in differential diagnostics, investigating this single characteristics does not provide information about the underlying deficit. Even though the acoustic correlates of stress were not different in the CAS and the PD group, the stress was still perceived as such.

2.1.3.3 Temporal properties

A more promising approach to reveal the core problem of CAS is investigation of speech timing. If CAS is in fact marked by a "...core impairment in planning and/or programming spatiotemporal parameters of movement sequences..." (American Speech-Language-Hearing Association 2007, p.4), it can be expected that temporal properties will be affected as well. A speaker who has problems realising appropriate articulatory movements will have affected also durations of all type speech events (e.g., segments, syllables and utterances).

The first group of studies addressing durations of segments in sequences with or without consonant clusters and with a different location of syllable boundary was done by Maassen, Nijland and colleagues. All participants in these studies were Dutch speakers.

Maassen et al. (2001) recorded five children with CAS and six typically developing children, aged 5;0 to 5;11, producing phrases with a target sequence /ə/#/sx/V (V CCV) or /əs/#/x/V (VC CV) (vowel was /a/, /i/ or /o/). The study demonstrated that a syllable boundary had a different effect on segment durations of typically developing children and children with CAS. Typical children significantly shortened /ə/ and lengthened /s/ in a sequence with consonantal clusters (V CCV) as compared to a sequence with single consonant (VC CV). Children with CAS, on the other hand, showed no differences in segment durations.

Similar results were obtained by Nijland et al. (2003b) who measured segmental durations in the speech of six children with CAS and six controls aged 4;11 - 6;10 years. Speech material consisted of sequences with a syllable boundary before the /sx/ cluster or with a syllable boundary separating the two consonants (/z/V1#/sx/V2/t/ vs. /z/V1/s/#/x/V2/t/; V1 was /ə/ or /ɔ/, and V2 was /a/, /i/, or /o/). Resulting syllables where either real or non-words. Durational measurements indicated that children with CAS produced longer segments (except V2) than typically developing children, and showed different effect of syllable structure. While typically developing children significantly shortened their clustered targets by shortening V1, second consonant in a cluster (/x/) and V2, children with CAS only shortened V2. Nijland et al. (2003b) argued that such behaviour causes impaired prosody in speakers with CAS since they are not able to adjust durations for the number of segments, or in unstressed compared to stressed syllables. Children with CAS were also less able to produce stable repetitions than control children suggesting less automated speech processes. However, the variability between speakers with CAS was not different from variability between typically developing children. Familiarity with the words did not have any influence on segment durations either for typically developing children or for CAS speakers.

Different aspects of evaluating speech durations in CAS was adopted by Shriberg et al. (2003b) who inspected the variability of duration by measuring speech and pause events in conversational samples of 15 children with sCAS, 30 typically developing children, and 30 children with speech delay of unknown origin. All participating children were between 3 and 6 years old. For each child a coefficient of variation was first calculated separately for speech and pause events, and the two results were used to calculate the coefficient of variation ratio between pause and speech events. Group results showed that children with sCAS had a lower coefficient of variation of speech events and higher coefficient of variation of pause events than the other two groups, resulting in higher ratio scores. Moreover, sCAS speakers showed a greater range of ratio scores than both the typically developing and the speech delayed children, and had a proportionally higher number of individuals scoring high values. The authors interpreted the findings of a reduced variability of speech events as evidence of a motor praxis and stress deficit observed in CAS but did not provide any explanation for the increased variability in pause events. I would like to suggest that this finding also reflects a problem in motor praxis. If every speech event demands a motor planning stage, which is impaired in CAS, the speakers could be expected to need a variable amount of time before commencing.

The importance of investigating temporal properties in CAS was pushed even further by Peter and Stoel-Gammon (2005) who explored the idea of a timing deficit as a core deficit in CAS. Two children with CAS aged 4;3 and 9;5 and two age matched typically developing children performed a series of speech and music tasks: sentence imitation (measured duration of vowel), non-word imitation (measured duration of vowel), monosyllabic word production (measured duration of onset, vowel and coda), singing (measured vowel onset interval), clapped rhythm imitation (measured clap onset interval), and paced repetitive tapping (measured tap onset interval). Compared to typically developing children, speakers with sCAS showed longer and more variable vowel durations on all three tasks measuring this parameter, longer coda durations, but at the same time similar onset durations. Speakers with sCAS were overall less accurate on the sentence and non-word imitations, with differences between the controls being higher for the non-words, and on all three music tasks. Comparing childrens' scores to adults' revealed that typically developing children were almost adult-like at singing, clapped rhythms and non-word imitations but not at sentence imitations. Speakers with sCAS achieved adult-like performance only on the clapped rhythm task. Although only two children participated in the study, Peter and Stoel-Gammon (2005) concluded that children with sCAS have less accurate speech and music timing than typically developing children, and that music timing is a good predictor of speech timing in CAS. Measurements of onset, vowel and coda durations allowed the authors to agree with Shriberg et al. (2003b) and Nijland et al. (2003b) that timing deficit is the underlying cause of observed stress impairments, since stress affects vowel and coda (both longer for sCAS speakers) but not onset (similar duration as control group).

The same authors further explored the potential of a timing deficit as a diagnostic marker of CAS by assessing timing abilities of 11 children (aged 4;7 - 6;6 years) with a moderate to severe speech disorder of unknown origin, and eleven age- and gender matched typically developing children (Peter and Stoel-Gammon, 2008). Children with speech disorders were additionally evaluated for the presence of CAS characteristics. The authors aimed to reveal how accuracy on speech tasks relates to accuracy on hand tasks, and how temporal accuracy is associated with the presence of CAS characteristics. All children performed a non-word imitation task, clapped rhythm imitation task, and paced repetitive task (children had to tap in time with a metronome and continue with the same pace after the metronome was switched off). Results confirmed that speech disorders of unknown origin are marked by a central timing deficit affecting speech and non-speech (hand) performance as evident from measures of timing accuracy on all tasks. They also showed that timing accuracy is inversely correlated with the number of CAS characteristics. Children with the highest accuracy scores had the least CAS characteristics and those with the lowest accuracy had the most CAS features.

Assessing speech and non-speech timing abilities seems promising as a diagnostic marker for CAS. Discovering impaired non-speech timing is particularly important since young children often cannot produce speech sequences needed for the evaluation of speech timing. On speech tasks, children with CAS showed longer segment durations, lack of temporal adjustments when segments are part of consonantal clusters, and no effect of syllable structure on the durations. Taken together, results on timing also give evidence for an impairment of motor control since durations of segments and syllables result from articulatory movements.

2.1.3.4 Articulatory movements

The last group of studies on speech production of CAS speakers has been addressing characteristics of speech motor control by assessing performance on diadochokinetic (DDK) task, F2 patterns and movements of articulators. Investigating articulatory movements seems to be the most appropriate approach to studying and revealing speech characteristics of CAS. As a motor speech impairment, movements of articulators can be expected to show deviant, and possibly unique, characteristics resulting from the core impairment.

Children with CAS have been commonly described as having problems in rapid articulatory changes, evident especially in poorer performance on the DDK task. The possibility of DDK as a diagnostic marker for CAS was evaluated by comparing three groups of Dutch speaking children: group of children with CAS (11 children: 6;3 - 7;9), dysarthria (nine children: 6;4 - 10;3) and typical development (11 children: 6;0 - 8;2) (Thoonen et al., 1999). All children had their maximum performance on DDK task, and on vowel and fricative prolongations assessed. Following this, validation groups of children with less clear diagnosis of CAS (ten children: 4;5 - 7;6) and dysarthria (nine children: 5;4 - 16;3), as well as typically developing children (11 children: 5;2 - 11;6) and those with non-specific speech disorders (11 children: 4;4 - 1-;11) were used to test the diagnostic power of the procedure. When comparing the diagnoses obtained by using this maximum performance task and those of speech therapists, they achieved 100% accuracy for CAS and 89% for dysarthria, with only one child being misdiagnosed. The analysis of individual measurements showed that CAS can be diagnosed on the basis of maximum rate of alternating sequences (/pataka/) and maximum fricative prolongation. The former clearly supports the notion of greater difficulty in sequencing speech movements for CAS (Thoonen et al., 1999).

Speech motor control was more extensively investigated in one study by Sussman et al. (2002) and a series of studies by Maassen, Nijland and colleagues. They explored the differences in coarticulation and the effect of syllable boundary by measuring F2 patterns in groups of children with CAS and control speakers.

Sussman et al. (2002) investigated production in five children with CAS (aged between 5;6 and 6;9 years) and three age-matched controls by analysing locus equations of CV syllables containing one of the stop consonants /b/, /d/, or /g/, and different vowels. Locus equations present linear regression of F2 from the consonant onset to the middle of the vowel (different vowels are included into the measure). Typically, slopes of the regression lines are the smallest for alveolars when compared to labials and velars, reflecting the least influence of the following vowel on the place of occlusion. Results from the typically developing child showed the expected patterns, while the speaker with CAS had an alveolar stop score very similar to the velar one, and higher than for labial, reflecting different places of occlusion in different vowel contexts for all three stops. This methodology also allowed the investigators to calculate the distance between the three stops in the acoustic space, revealing that CAS speakers have more closely spaced /b/, /d/, and /g/ categories than the control group.

Maassen et al. (2001) recorded five Dutch children with CAS and six control children, aged 5;0 to 5;11, producing phrases with a target sequence /ə/#/sx/V (V CCV) or /əs/#/x/V (VC CV) (vowel was /a/, /i/ or /o/). Although the results showed no influence of target sequence on F2 trajectory in any of the speaker groups, it did reveal a different vowel type effect on the anticipatory correlation. Children in the control group showed the effect of the upcoming vowel only on /x/, while in the speech of children with CAS the vowel had influenced both consonants. Anticipatory articulation was thus extended further in the CAS group.

The same was observed in a subsequent study by Nijland et al. (2002). They were tracing F2 patterns in repetitions of non-word /ə/CV sequences (consonants /s/, /x/, /b/, and /d/, and vowels /i/, /a/, and /u/) of nine children with CAS (5;0 - 6;10 years), six typically developing children (4;9 - 5;11) and six adults (20 - 30 years). All participants spoke Dutch. F2 values were measured at six points along the sequence: mid /ə/, end /ə/, consonant, onset of transition, end of transition, and mid vowel. The measurement firstly showed that CAS children have the highest average F2 values, and adult women the lowest of the three groups. Additionally, they also had higher within-speaker variability than control children who in turn had higher within-speaker variability than adults. An interesting observation was however, that between-speaker variance inside the group did not significantly differ for the three groups. Speakers in each of the groups performed equally variably. To investigate the distinction between sequence types Nijland et al. (2002) used a ratio between each speakers averaged F2 values for a sequence with the vowel /i/ and a sequence with the vowel /u/. The higher the ratio, the bigger the difference in production of the two sequences. This measure revealed that adults and typical children displayed similar patterns across the sequences, while CAS speakers did not. CAS speakers also showed less distinction in F2 at vowel mid point than the other two groups implying a poorer ability to discriminate between vowels. Inspecting values at the six points of measurements allowed the authors to observe anticipatory intersyllabic coarticulation of the final vowel on the initial /ə/, and anticipatory intrasyllabic coarticulation of the final vowel on the preceding consonant. Typically developing children and adults again showed a consistent coarticulation effect on both points of measurements while children with CAS did not. They displayed large intra- and intersyllabic coarticulation for some consonants and not for others and differed from each other over the different sequences.

The above study was further extended to incorporate the effect of syllable boundary on F2 patterns (Nijland et al., 2003b). This time six Dutch speaking children with CAS and six controls aged 4;11 - 6;10 took part. As in previous studies by Nijland and colleagues, speech material consisted of sequences with a syllable boundary before the /sx/ cluster or with a syllable boundary separating the two consonants (/z/V1#/sx/V2/t/ vs. /z/V1/s/#/x/V2/t/; V1 was /ə/ or /ɔ/, and V2 was /a/, /i/, or /o/). The syllables were additionally controlled for their frequency and divided into high frequency and low frequency groups. Measurements revealed that a significant effect of syllable structure was evident in the F2 patterns and in the ratios between sequences with /i/ and /o/ as V2. Both groups of children, but especially CAS speakers, had significantly higher F2 values at mid V1, onset of the /x/-V2 transition and at mid V2 in a sequence with a cluster than in the /s/#/x/ one. Inspecting i/o ratios at the first vowel showed that inter-syllabic coarticulation was stronger for the sequence with a cluster, as the ratio for mid V1 was significantly different in the two structure types. Syllable structure had, however, no effect on the intra-syllabic coarticulation as measured by comparing the effect of V2 on /s/ in both conditions. But the two groups did differ in the extent of anticipatory coarticulation. Just as in the study by Maassen et al. (2001), influence of V2 was observed already at /s/ in the CAS group, and influence on /x/ was higher for the CAS than for the control speakers.

The motor origin of CAS was further tested by observing speakers' adaptation to bite-block condition (Nijland et al., 2003a). Five children with CAS and five typically developing children aged 5;0 - 6;10, as well as six young women participated in the study. Just as in previous studies (Nijland et al., 2002), all participants were Dutch speakers and all made repetitions of /ə/CV sequences in which F2 was measured at six different points. The experiment clearly showed that speakers with CAS compensate differently to bite-block, meaning that they adapt motor plans in a different way. Although typically developing children were not able to compensate to the same extent as adults as evident from higher F2 values, the F2 ratio between sequences with /i/ and /u/ did not differ for these two groups. Children with CAS, however, again showed different F2 values over the sequence. They showed a bite-block effect already at the end of /ə/ as compared to transition onset for adults. The effect was observed in a bigger number of possible CV combinations for CAS speakers than for control children. F2 ratios showed that bite-block actually improved vowel production for the CAS group but overall the pattern of ratio over the sequence was different than for control children and adults. An expected result was that variability was higher for children with CAS in both conditions.

Only two studies investigating motor control abilities of speakers with CAS by observing tongue movements have been published so far. One used electropalatography (EPG) and the other electromagnetic articulography (EMA). Both techniques present a rich source of information about the characteristics of tongue movements. They both allow observation of place of articulation (tongue palate contact only in the case of EPG) and the timing of movements. EMA additionally provides information about tongue position and shape, relations between different parts of the tongue and the relation of its movements to the movements of other articulators. The last two measures depend on the number of coils attached to the tongue and to other articulators. Articulatory data is usually recorded at the same time as acoustic one, with the two signals being temporally aligned. In this way, recorded speech can be easily segmented into target segments or sequences, and both articulatory and acoustic characteristics can be extracted and analysed. Because CAS is a motor speech disorder, it seems very reasonable to explore it with articulatory studies. Of particular interest is investigation of tongue movements in speech. The tongue is one of the busiest articulators and it can be expected to show different movement characteristics and patterns which are the cause of a number of observed features of CAS (e.g., segmental errors, inconsistency in speech production, groping movements). Direct observation of tongue movements is also expected to provide richer and more reliable data than conclusions made from the acoustic signal and measuring formant values. Such reasoning was the main motivation for the studies presented below, and for the study presented in this thesis.

The EPG study by Barry (1993) investigated place of articulation in different groups of speakers with a speech motor disorder and three adults (30 - 50 years). The experimental groups included two speakers with CAS (both 9 years), two speakers with developmental dysarthria, one speaker with acquired dyspraxia and one with acquired dysarthria. All speakers uttered words with simple syllable structures CV or CVC, where the consonant was either /d/, /k/, /s/, /z/, /ʃ/, or /l/. As could be expected, performance of the two children with CAS were different from each other, since CAS is marked by high intra- and inter-speaker variability. One of the speakers realised all stops and most of the fricatives as velars and completely omitted /l/. The tongue-palate contact was longer for target alveolars, and shorter for target velars and fricatives than observed in adults. This speaker also had higher variability in target alveolars than the control group but similar variability for velars. In contrast, the other CAS child could articulate all the sounds at the correct place and with durations similar to adults', except the alveolar fricatives which were realised as velar stops and shorter than in adults. The author concludes that as a group, apraxic speakers (two speakers with CAS and one with acquired apraxia) were less able to control placement and timing, supporting the theory of a motor deficit in apraxia of speech.

In the only EMA study of CAS, Nijland et al. (2004) investigated coordination of gestures in three Dutch speaking children with CAS (4;3, 9;0, and 11;10 years) and three typically developing children (9;8, 10;3, and 12;11 years). After four coils were attached to the participants' tongue body, tongue tip, upper lip and lower lip, children made sequences of CV, VC, C1C2V, and C2C1V syllables. The following sequences were selected and analysed by using relative phase values, capturing movement of the upper lip with regard to lower lip, and of tongue tip with regard to lip opening: /pa:/, /a:p/, /spa:/ (a common cluster in Dutch), /pa:s/, /ta:p/, and /pta:/ (an unusual cluster in Dutch). The first difference between the groups was observed already during the recording. While typically developing children could produce all necessary sequences, the youngest child with CAS could only produce CV and VC syllable structure, and the other two children with CAS could not produce /pt/ sequences without inserting a vowel (one speaker was inserting /a:/ and the other /ə/). Measurements revealed even more differences between the groups. Control children achieved more similar scores on each of the targets than did CAS speakers. CAS speakers were additionally marked by higher variability of measurements. The groups also showed a different effect of the familiarity of a cluster on the production. Typically developing children had similar relative phase scores for common clusters and different for unusual one. Children with

CAS did not show such an effect and had similar scores also for unusual clusters. The authors concluded that different motor patterns resulted from coordination problems of CAS speakers.

The group of studies focusing on speech motor control in CAS suggest a deviant performance of these speakers. Children with CAS show poorer abilities of repeating /pataka/ sequences, lack of ability to adapt articulatory movements to syllable boundary, more extensive anticipatory coarticulation, different adaptation to bite-block, more errors in place of articulation and more problems coordinating upper lip, lower lip, and tongue tip than typically developing children and adults. This also supports the necessity for more articulatory studies of CAS speech. Direct observation of movements of articulators could reveal a diagnostic marker and non-invasive articulatory techniques could become part of diagnostic procedure.

2.1.3.5 Relation between speech perception and production

Speech perception in CAS has received much less attention in research than production. This may be mainly because speakers with CAS have always been described as having better perceptive than expressive language skills. The focus of the research was thus on the latter. However, a few studies that did explore speech perception in CAS show the existence of perceptual impairments as well. Even more, they revealed similar problems in perception and in production. Although the focus of this thesis is on the tongue movements in CAS and it does not include investigation of their perceptual skills, I believe it is important to present perception studies as well. In that way, this section will provide the reader with different understandings of CAS and present the complexity of disorder better.

Two studies investigating the relation between perception and production skills in CAS and typically developing children have been done so far. In the first one, Groenen et al. (1996) evaluated 17 CAS speakers and 16 controls aged between 6;11 and 11;6 years. Children were first assessed on identification and discrimination tasks of monosyllabic CV words differing in the voiced stop consonant. Participants had to assign targets along a seven step /b/-/d/ continuum into a consonant group, and discriminate between pairs of the same targets. Plotting responses of the identification task revealed that children in both groups have a similar slope of identification curve which according to the authors suggest equally consistent phonetic processing. The main difference between the groups was that CAS speakers had their phoneme boundary shifted towards the alveolar target. Alveolar place of articulation was perceived even for the targets that were more similar to /b/ than to /d/. Even more differences between the groups were evident on the discrimination task. CAS speakers performed more poorly than controls and were less able to discriminate between the pairs of targets. The authors concluded that children with CAS have poorer auditory processing and poorer access to information in auditory memory. In addition to perception, participating children were assessed also by two articulation tests of words and non-words. Results showed that disturbances in the discrimination task were closely related to the number of place of articulation substitutions in production, but not to the manner of articulation or voicing. The same feature, place of articulation, was thus impaired in production and in perception of CAS speakers.

Problems with identification of consonants were evident also in a study of three CAS children aged 8, 11 and 12 years, and a typically developing child aged 9 years (Sussman et al., 2002). They were asked to assign items from a 14-step CV /b/-/d/, and /d/-/g/ continuum into one of the three stop categories. The typically developing child showed an expected pattern of sound classification with a sharp phoneme boundary for both labio-alveolar, and velar-alveolar pairs. Performance of children with CAS was different. Two of them showed normal labial-alveolar boundary and deviant velar-alveolar boundary while the third child showed the exact opposite performance. Relating observation about speech perception to the production part of the study (described earlier in section 2.1.3.4) showed that children with CAS had impaired control of (stop) place of articulation both in production and in perception.

Children with CAS do not have problems only with the perception of individual segments, but also with perception of syllables. In a study by Marquardt et al. (2002) three children with CAS and three control speakers (6 - 8 years old) were asked to identify the number of syllables in one- to four-syllable words, the location of difference in minimal pairs of CVC words, and the position of consonants in CV, VC, CVC, CCVC, CVCC, CCV, and CCCVC syllables. Children with CAS showed poorer performance on all three tasks than control children of similar age. They made more errors in identifying the number of syllables, particularly in the three- and four-syllable words, were less able to identify the position of difference in minimal pairs and made errors in all three possible positions, and had more trouble identifying consonant position, particularly in complex syllable structures. As noted in all other studies on CAS, individual speakers with CAS performed differently on the tasks. For example one of the children with CAS performed almost at the 100% correct level in identifying the number of syllables in words and was also better at the remaining two tasks. However, the three participants with CAS showed evidence of a disrupted ability to perceive syllables and to compare them in terms of minimal pairs and different structures. Marquardt et al. (2002) also proposed that the disruption at the syllable level should affect suprasegmental elements that are related to the syllable, and would explain the reported impairment of prosody in CAS. Again, this can be related to findings about syllable production in CAS which showed that speakers have more problems with longer sequences, consonant clusters, and lack of ability to adjust productions according to the syllable boundary.

Investigating perception and production abilities of speakers with CAS was taken a step further by Nijland (2009) who tested the idea that children with higher-level PD show higher-level perceptual problems, while children with lower-level speech motor disorders like CAS have lower-level perceptual problems. This study included two groups of Dutch speaking children, 21 children with speech disorders diagnosed as either CAS (six children), PD (four children), or mixed (11 children), and a control group of 20 typically developing children. All participants were between 5;5 and 7;11 years old. Children first underwent a series of speech production tasks recording percentage of correct consonants and vowels, and percentage of consonant substitutions. They also participated in five auditory tasks assessing their ability to discriminate between non-word pairs, to judge whether the words in pairs rhyme or not, and to identify a different tone in pairs of 3-tone patterns. They were tested on categorical classification and discrimination tasks with a seven-step continuum from /pop/ to /kop/. The tone task was expected to give information about the children's auditory temporal processing, and the remaining tasks about their higher-level perception (rhyming task and categorical classification task) and lower-level perception (non-word discrimination task and categorical discrimination). Children were also assessed on memory tasks including hand movements, number recall, word order, and spatial memory.

Comparing the results of the non-word discrimination and rhyming tasks across the groups revealed that typically developing children performed equally well on both of them. The same was not observed in the other three groups. Overall, children with speech disorder achieved lower scores on both tasks and also had lower scores at the rhyming than at the non-word task. However, the results of the non-word task were significant only between the CAS and control groups, while the results of the rhyming task differentiated both CAS and PD groups from the control one but not from each other. A group of children with mixed disorders did not differ from any other group. The difference between the performance on the higher-level and lower-level perception tasks was also significant only in the CAS and PD groups. A categorical classification task showed that although all children showed similar classification patterns they did differ significantly in the percentage of responses at the phoneme boundary and in end-points of the curves. They also performed similarly on the categorical discrimination task, showing no significant difference between groups, although speakers with speech disorder showed poorer ability to discriminate stimuli from different categories.

The memory task revealed that children with CAS had lower scores than typically developing children on number recall, word order and spatial memory task but similar scores on hand movement task. In contrast, children from the mixed group had lower scores on number recall and word tasks, and children with PD only on the word task.

Inspecting the correlation between production and perception data showed that children who were better at the real-word production task did better at the discrimination and rhyming tasks as well, and that higher number of errors in manner and place of articulation were related to poorer performance on the rhyming task. In the production of non-words, percentage of correct consonants and vowels were significantly correlated to the performance at discrimination and rhyming tasks, while errors in manner of articulation were negatively correlated to the same two tasks. Overall, results showed a connection between production and perception in different speech disorders. Although children with CAS did not show only lower-level problems, children with PD did show problems only on the higher-level perception task. The author suggested that higher-level perception problems in CAS are due to development but in my opinion they might be a result of impaired lower-level perception skills.

The last ability to be tested in this study was tone discrimination. Interestingly, this task proved to be too difficult for children with speech disorders since only two out of 21 could perform, compared to 15 out of 20 typically developing children. Nijland (2009) concluded that this inability shows a deficient memory, although they did not perform equally badly on all memory tasks. Children in PD group had problems only on word task, mixed group additionally on number recall, and children with CAS even with spatial memory.

Findings about similarly impaired speech production and perception in CAS contribute enormously to the knowledge about the disorder and about necessary therapy. First of all, they give evidence for the disruption of the whole speech processing system, and open questions about primary and secondary deficits. Nijland (2009) presented the idea that the core impairment lies in speech perception and that production becomes impaired via the motor mirror neurones which get activated in perception and cause wrong speech production processing. Alternatively, impaired perception could be a direct result of poorer and inconsistent production (Raaymakers and Crul 1988 in Nijland 2009, p.224). Secondly, the evidence of similarly impaired production and perception implies that both aspects of speech have to be addressed in speech therapy to improve not only perception of therapist's speech but also auditory feedback of speaker with CAS.

2.1.4 Summary

To summarise, despite all the problems with selecting children who have CAS and not some other kind of speech disorder, participants in the studies presented in this chapter showed similar speech profiles. They produce vowel and consonant errors, impaired stress patterns, deviant segment and syllable durations, different F2 patterns, impaired coordination of articulators, problems with speech perception, different effects of word familiarity, syllable structure and syllable boundary location, and high within- and between-speaker variability. The review also showed that CAS cannot be reliably diagnosed only on the basis of the number and type of segmental errors or auditory assessment of stress. It has to be evaluated by assessing the ability to express and control different acoustic and articulatory features of speech. Articulatory assessment is particularly useful because it allows direct observation of articulators during speech and combining articulatory with acoustic information. Addressing temporal and articulatory features is especially valid also because it targets those speech characteristics that are directly affected by problems with spatiotemporal planning of movement sequences. In this way, main and differentiating characteristics of CAS could be more easily uncovered. Another important conclusion based on the review of earlier studies of CAS, is that the disorder seems to be even more complex than thought, with speakers having affected not only speech production but in a similar way speech perception as well. Such observation will probably cause another switch in our understanding of the disorder.

The research presented in this thesis is, however, focused on the temporal and articulatory characteristics of CAS speech as both types of information can be obtained by ultrasound recording. In order to uncover those characteristics that define CAS, the same characteristics have to be investigated in the control groups of speakers without any speech and language impairments. In the research presented in the later chapters, speech of speakers with CAS was compared to a control group of adult speakers and a control group of typically developing children. Selecting these two control groups enabled comparing speakers with CAS (all teenagers in this case) not only with standard adult productions but also with immature productions of typically developing children. Having two previously well researched groups additionally served as a validation of the applied methodology. The next two sections in this chapter, 2.2 and 2.3, describe previous research findings on temporal properties and tongue movements of the two age groups.

2.2 Temporal properties in speech of adults and typically developing children

Speech is one of the most complex and fast human motor activities. It is typically produced at the rate of about 15 speech sounds per second (Levelt, 1989) or three to six syllables per second (Crystal and House, 1982) but it can be as fast as nine syllables per second (Kent, 2000). Although most commonly English syllable types are simple combinations of one consonant and one vowel, they can be composed of combinations of up to three consonants which articulatorily present very complex targets. In order to produce intelligible output, a speaker has to be able to apply correct durations to the individual segments and to the whole syllables. As shown in a number of studies presented in this sections, adults are able to systematically adjust segment and syllable durations depending on the complexity of the speech material. Typically developing children develop the same ability through time but typically need more time for the correct realisation of the targets.

2.2.1 Adults

Duration of individual speech sounds and whole syllables in adult speech have been investigated in great detail. They are relatively easy to measure from the acoustic signal and reveal important information about the nature of speech sounds, and the effect of their combinations and syllable structure.

2.2.1.1 Duration of individual speech sounds

Several studies have focused on the durations of individual segments (Klatt, 1976; Umeda, 1977; Farnetani and Kori, 1986). They mostly agree that both vowels and consonants are affected by the type of the segment (reflecting its intrinsic duration), by the stress which prolongs both vowels and preceding consonants, by the adjacent segments resulting in shorter segment durations in consonantal clusters than as singletons and in shortening of the vowel with an increase of the coda segments, and by the position in the syllable, with coda consonants being longer than onset consonants.

The study employed in the research reported later in this thesis focuses only on the duration of syllables with onsets of different complexity and constituent segments. The consonants were /p/, /s/, and /l/. Because of that, some previous findings about the duration of these three consonants are summarised below.

Word initial English /s/ has been reported to last from 113 ms to 155 ms (Haggard, 1973; O'Shaughnessy, 1974; Klatt, 1974; Umeda, 1977; Crystal and House, 1988; Greenberg et al., 2003), duration of word-initial /p/ ranges between 89 ms to 117 ms (Umeda, 1977; O'Shaughnessy, 1974; Crystal and House, 1988; Greenberg et al., 2003) and, finally, duration of a word-initial /l/ was measured between 65 ms and 100 ms (Haggard, 1973; O'Shaughnessy, 1974; Umeda, 1977; Crystal and House, 1988; Greenberg et al., 2003). As a syllable-initial singleton, /s/ is the longest of the three consonants, followed by /p/, and /l/, which is the shortest.

However, when these segments are part of an onset cluster they each change in their own specific way, depending on the adjacent consonants. Initial English /s/ is shortened to 76 - 86 ms (Klatt, 1974; O'Shaughnessy, 1974; Umeda, 1977) when followed by a stop. More variation in the literature has been reported for the duration of /s/ preceding /l/. Cluster /s/ can become longer (Greenberg et al., 2003), equal (Crystal and House, 1988, 1990; Haggard, 1973) or shorter (O'Shaughnessy, 1974).

English /p/ in an initial onset position lengthens when followed by a voiced consonant (O'Shaughnessy, 1974; Umeda, 1977), and shortens as a second segment in a cluster (O'Shaughnessy, 1974; Klatt, 1974).

In contrast to the other two consonants, /l/ cannot be the first segment of a consonant cluster in English. In the second position of a cluster /l/ shortens when preceded by a fricative (Haggard, 1973; O'Shaughnessy, 1974; Umeda, 1977; Crystal and House, 1988) and either shortens (O'Shaughnessy, 1974) or stays the same when it follows an unvoiced stop (Umeda, 1977).

These studies clearly show the effect of onset structure on the constituents' duration but they do not completely agree on the exact type of the effect. Differences between the studies could be due to different speech materials, numbers of speakers and speaking rate. Haggard (1973), Klatt (1974) and Crystal and House (1988, 1990) measured segment durations in words read in isolation (eight, three, and six participants, respectively), O'Shaughnessy (1974) in words spoken in a sentence (one speaker), Umeda (1977) in a read passage (one speaker), and Greenberg et al. (2003) in dialogues (581 participants).

The authors of the studies presented in this section additionally tried to explain the source of durational differences between segments, particularly between singleton and cluster realisations. O'Shaughnessy (1974) stated that shorter durations of segments in a cluster result from shorter distances the articulators have to travel in the realisation of a cluster. Umeda (1977), on the other hand, relates the duration of a consonant to the articulator and/or type of gesture shared between the consonant and its adjacent consonant. Duration of the consonant is different when the gesture is overlapping than when it is conflicting. Sharing a gesture additionally prevents the consonant from shortening, and sharing an articulator (e.g., the tongue in /st/) lowers the variance of timing.

Duration of a segment in a consonantal cluster thus always depends on its articulatory realisation. For this reason, speakers with CAS could be expected not to show any clear patterns of adapting segment durations in clusters. Because of their problems in planning articulatory movements and because of the inconsistency in their realisations, they would produce different durations on different attempts of the same targets. The temporal properties of clustered segments would differ from those of adults.

2.2.1.2 Syllable duration

Studies investigating temporal properties of syllables of different complexity revealed that syllable duration is mainly affected by the syllable's stress, by the number of segments in a syllable and by the type of the syllable's segments with some additional influence of the number of syllables in a word, the position of a syllable in a word and speech rate (Farnetani and Kori, 1986; Crystal and House, 1990; Greenberg et al., 2003; Meyer, 1994).

Syllable duration was investigated in great detail in two studies by Crystal and House (1990) and Greenberg et al. (2003). The first study was based on six English speakers reading two short scripts and the second on SWITCHBOARD, a corpus of spontaneous American English dialogues, with both data sets containing mainly monosyllabic words. Both studies were looking at the durational characteristics of syllables differing in the number of onset and coda segments and both make a distinction between stressed and unstressed syllables. Crystal and House (1990) additionally separate syllables into prepausal and not prepausal groups.

Table 2.1 presents average syllable, vowel and consonant durations for both stressed and unstressed conditions and for different syllable structures, as measured in these two studies. As can be seen, syllable duration increases with increasing number of segments in both stressed and unstressed conditions. Stressed syllables and their segments have greater duration than their unstressed counterparts, and non-prepausal syllables are shorter then prepausal ones.

Crystal and House (1990) concluded that the duration of a syllable depends on its stress condition and is highly correlated with the number of segments. However, the changes did not affect all parts of the syllable in the same way. The authors focused only on the vocalic segments and observed that the changes in vowel duration depend only on the stress and not on the number or the order of the syllable segments. As syllable complexity increases, the nucleus of the stressed syllable remains practically stable while the nucleus of the unstressed syllable tends to shorten. Although not reported, inspecting the presented data points to a conclusion that the same seems to be the case for the consonantal segments which are additionally shorter in the second position of an onset cluster, both in stressed and unstressed condition. Durations of segments are moreover influenced by the presence of a pause following the syllable. Both vowels and coda consonant are lengthened in prepausal syllables. In accordance with the finding of increasing syllable duration with the increasing number of segments, they concluded that syllables do not have intrinsic durations. Durations of segments within the syllables in their data did not undergo any adjustment in order to achieve a constant syllable duration. It is however necessary to mention again that this observation was made on the basis of all syllable types, without controlling for syllable complexity. The authors also suggested that it might be due to the read speech material used in the study.

The same general conclusions were made in the study by Greenberg et al. (2003). As can be seen in Table 2.1 syllable duration increases with the addition of segments, both in the onset and in the offset, and it is greater in the stressed than unstressed condition. However, stress influence additionally depends on the syllable's position in the word and the segment's position in the syllable resulting in different parts of syllable being affected in a different way. Overall, the stress condition has a greater impact on the vocalic part of the syllable. In a word-initial syllable, the duration of a vowel changes with the addition of an onset or an offset, and the average durations of an onset's segments stay stable no matter how the syllable complexity changes both

	Crystal and House (1990)												Greenberg et al. (2003)							
	not prepausal (ms)						prepausal (ms)						(ms)							
Syllable	syllable		V		C		syllable		V		C		syllable		V		onset C		offset C	
structure	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
V	130	70	130	70			243	89	243	89			154	75	154	75				
VC	199	120	127	64	72	56	282	173	197	95	85	78	258	142	172	72			86	70
VCC	246	147	133	60			337		135											
CV	206	101	131	55	75	46	350	134	273	65	77	69	231	124	135	64	96	60		
CVC	280	156	129	55			377	190	189	72			310	184	139	56	95	63	85	65
CVCC	353	269	126	52			457	378	149	84			400	244	126	55	101	62	86	70
CVCCC	400		113				426		104											
CCV	281		133				513		308											
CCVC	358	213	122	39			448	273	166	67			382	249	134	58	86	61	86	69
CCVCC	407		128				463		135											
CCCVC	483		104				516		139											

Table 2.1: Average syllable, vowel and consonant durations (ms) for different types of syllable structures and stress conditions. (+) marks stressed syllable and (-) unstressed syllable, as reported by Crystal and House (1990) and Greenberg et al. (2003).

in stressed and unstressed syllables. Onset segments in stressed syllables are, though, 41% to 63% longer than their unstressed counterparts.

2.2.1.3 Summary

These previous works clearly demonstrate that although the syllable's duration increases with the number of syllable segment, the increase itself it is not linear or the sum of individual segments. Individual segments have their own intrinsic durations when produced as singletons in isolation or in simple consonant-vowel combinations. However, speech segments are only rarely produced in isolation. They are almost always part of a continuous speech in which segments affect each others production. When producing continuous speech, segments influence each other by adjusting optimal trajectories of individual segments in a way which allows the best balance between the demands of the correct output and the physiological demands of the articulatory system. The latter tries to run as smoothly and efficiently as possible (Lindblom, 1990). Articulatory gestures necessary for the production of connected segments are thus adapted regarding the demands of the whole sequence. This is especially visible in the case of consonant clusters where two or more transition phases are executed in a sequence.

Overall, the above studies showed that adult speakers systematically increase syllable durations with the increased number of syllable segments, either in the onset or in the offset. Speakers with CAS would be, again, expected no to show such straightforward adaptation. Their impaired planning is expected to affect syllable level as well. Because of problems with the execution of individual segments, and even more problems with clusters, they would produce different, inappropriate syllable durations, masking any effect of the number of syllable segments.

2.2.2 Typically developing children

Speech characteristics of typically developing children reflect the ongoing process of speech development. Because of the complexity of speech production, speech needs several years of development to achieve adult like functioning. The development depends on the maturation of the motor, cognitive and linguistic system. During the speech development process speech of children differs from speech of adults.

One of the most detailed and still most commonly cited papers on the topic of acoustic characteristics of developing speech was written by Kent in 1976 (Kent, 1976).

In this paper the author reviews existing literature showing evidence of a decrease of mean values and variability of several acoustic parameters as an effect of increasing age. He reports measured decline in fundamental frequency between one and 12 years of age, decrease in variability of F1 and F2 which end at adult-like levels at 11 or 12 years of age (Eguchi and Hirsh 1969, cited in Kent 1976, p.433), systematic changes in VOT during the first six years of life and achieving adult-like values by eight years of age, decrease in variability of duration with increasing age (DiSimoni 1974, cited in Kent 1976, p.438) and achieving adult-like levels of relative variance in speech timing by 11 years (Tingley and Allen 1975, cited in Kent 1976, p.439).

The main conclusion of the review is that adult-like accuracy of speech motor control is achieved around the onset of adolescence, by 11 or 12 years of age. However, it has to be pointed out that this conclusion is based on studies investigating speech in children up until 13 years of age only. The selection of upper age limit was heavily influenced by the idea that speech maturation is linked to the maturation of the neural system, particularly to the process of neuron myelinasation which ends around the onset of puberty and thus older children were not included into these studies. As can be seen from later research including older teenagers, temporal characteristics of speech are not completely matured and adult-like at the onset of adolescence but do keep stabilising during the teenage years. Nonetheless, Kent's review still provides strong evidence for increasing timing control with age.

2.2.2.1 Acquisition of speech sounds and consonantal clusters

Before describing temporal, and in a later section articulatory, properties of children's speech any further it is important to understand the time line of typical speech acquisition. This section is a brief presentation of ages at which individual segments and consonantal clusters are acquired.

Acquisition of speech sounds and clusters is a continuing process spanning from birth until up to nine years of age. Table 2.2 presents ages at which 75% of tested children correctly produced individual segments in initial and final word position in a study by Templin (1957) and 90% of children in a study by Smit et al. (1990). Table 2.3 presents ages for word-initial consonant clusters in the same studies. As this data shows, children acquire correct, adult-like productions at different ages, but they all follow a similar developing path with some speech sounds acquired before the others and with individual segments mastered before clusters, although at the same time intelligibility of speech is not significantly reduced.

consonant	Templin (1957)	Smit et al. (1990): females	Smit et al. (1990): males
/m/	3;0	3;0	3;0
/n/	3;0	6;6	3;0
/ŋ/	3;0	7;0-9;0	7;0-9;0
/h/	3;0	3;0	3;0
/w/	3;0	3;0	3;0
/j/	3;6	4;0	5;0
/p/	3;0	3;0	3;0
/b/	4;0	3;0	3;0
/t/	3;0	4;0	3;6
/d/	4;0	3;0	3;6
/k/	4;0	3;6	3;6
/g/	4;0	3;6	4;0
/f/	3;0	3;6	3;6
/v/	6;0	5;6	5;6
/θ/	6;0	6;0	8;0
/ð/	7;0	4;6	7;0
/s/	4;6	7;0-9;0	7;0-9;0
/z/	7;0	7;0-9;0	7;0-9;0
/ʃ/	4;0	6;0	7;0
/tʃ/	4;6	6;0	7;0
/dʒ/	7;0	6;0	7;0
/l/	6;0	6;0	7;0
/r/	4;0	8;0	8;0

Table 2.2: Ages of acquisition of speech sounds in initial and final position as presented in Templin (1957) (75% criterion) and Smit et al. (1990) (90% criterion).

consonantal cluster	Templin (1957)	Smit et al. (1990): females	Smit et al. (1990): males
/tw, kw/	4;0	4;0	5;6
/sp, st, sk/	4;0	7;0-9;0	7;0-9;0
/sm, sn/	4;0	7;0-9;0	7;0-9;0
/sw/	7;0	7;0-9;0	7;0-9;0
/sl/	7;0	7;0-9;0	7;0-9;0
/pl, bl, kl, gl, fl/	4;0-5;0	5;6	6;0
/pr, br, tr, dr, kr, gr, fr/	4;0-4;6	8;0	8;0
/θr/	7;0	9;0	9;0
/skw/	6;0	7;0-9;0	7;0-9;0
/spl/	7;0	7;0-9;0	7;0-9;0
/spr, str, skr/	5;0-7;0	7;0-9;0	7;0-9;0

Table 2.3: Ages of acquisition of word-initial onset clusters as presented by Templin (1957) (75% criterion) and Smit et al. (1990) (90% criterion).

On average, the first acquired speech sounds are sonorants and stops, followed by fricatives, laterals, and /r/-sounds. This is reflected in clusters as well. The first clusters are combinations of stops and sonorants, followed by clusters containing /s/ and /θ/. Two consonants clusters are additionally mastered before the three consonants ones which are acquired by 75% children around six years of age and by 90% between seven and nine years. However, it has to be pointed out that these acquisition ages offer only guidelines because of the great variability observed among children. Additionally, the 75% criterion implies that one out of four children still does not produce the correct target. For that reason, a 90% criterion provides a better assessment of the age of speech sound acquisition.

Based on the reported data it can be concluded that most children achieve adult-like production of most segments and clusters by nine years of age. Any errors observed at a later age potentially indicate presence of a speech disorder.

2.2.2.2 Duration of individual speech sounds and syllables

A fairly consistent observation about the temporal characteristics of children's speech is that children have longer segment durations and greater variability of segment duration than adults and that both parameters decline with increasing age until adult-like values are achieved with the maturation of the speech motor control system (DiSimoni, 1974a,b; Gilbert and Purves, 1977; Smith, 1978; Kent and Forner, 1980; Smith and Kenney, 1998; Lee et al., 1999).

Smith (1978) investigated temporal properties of nine non-word stop-vowel combinations, CVC(VC), in ten children aged 2;9 to 2;11, ten children aged 4;1 to 4;7 and ten adults (23 to 40 years of age). The results clearly showed a decrease in duration with increasing age for both word and segment durations. Younger children had average word durations 38-122 ms longer than older children and 97-131 ms longer than adults, and older children 40-136 ms longer than adults. The measurements of individual segment durations revealed similar results; younger children had on average about 50 ms and older children about 25 ms longer segment durations than adults. Additionally, the author reports that 2-year-olds revealed the greatest variability in word durations (standard deviation (S.D.) 47-146 ms) while 4-year-olds had average variability over all nine words similar to those of adults (S.D. 44-65 ms and 44-68 ms, respectively) with even less variability than adults for some of the targets.

The same trend of decreasing duration of segments with increasing age was further established in a study by Kent and Forner (1980) who investigated ten speakers in four age groups: 4-year-olds, 6-year-olds, 12-year-olds and young adults. Participants in this study made repetitions of three different sentences and the authors measured phrase, word, vowel, consonant, VOT and closure durations for selected targets. The results showed a decrease in duration for all measures and in their variability with age. However just as in the study of Smith (1978) even some of the 4-year-olds had a similar S.D. to adults, implying a great variability in the motor control domain between speakers of the same age.

Because of the great inter-speaker variability observed in speech timing it is interesting to follow the development of the same speakers over time. Smith and Kenney (1998) have studied speech production of seven girls, each recorded three times at 8-9, 9;6-10;6 and 11-12 years of age, and seven young adult females (22-25 years). The general conclusion was that all measured acoustic parameters were approaching adult-like values over time. Duration was measured for the syllable /ɪs/ in a word "sissy". The average values for all seven girls were 294, 253, and 240 ms at the three different ages and 203 ms for the adults, showing a linear decrease with age. Looking at each individual speaker's changes in duration reveals that two out of the seven girls had durations at or very near the adult values already at the first recording; four more girls reached this target at the second recording, with the last one joining them at the third. Another important result of this study was the finding that even 11-12 year-olds do not exhibit the same durations as adults since the difference between the two groups was still about 15%.

In one of the biggest studies of the acoustics of children's speech Lee et al. (1999) investigated 436 children aged 5-18 years (every year group has subgroups of 9-25 male and female participants) and 56 adults. The recorded material consisted of ten monophthongal and five diphthongal words and five phonetically rich sentences. Duration was measured for all sentences, vowels and the consonant /s/. All three measurements showed shortening in durations and reduction of within- and between-speaker variability with an increase in age. Vowels were the longest at five and six years of age (279 and 264 ms, respectively) and significantly longer than at all the other ages. Furthermore, significant decrease was observed from 10 years (199 ms) to 12 years (178 ms) and from 11 (191 ms) to 15 (168 ms), but there was a significant increase from 15 year-olds to adult values (approximately 185 ms). A significant drop in within-subject variability was noted only between 11 and 13 years of age suggesting that speakers stabilise the duration of their vowels and reach adult-like values at around this age. Unfortunately, the authors do not report at what time between-subject variability becomes adult-like. The chart only suggests that the least variability was observed at 15 years but there is no information on whether it was significantly different from the adults or not. Very similar observations were made for duration of /s/ and entire sentences. /s/ significantly decreased between the age of 10 (170 ms) and 12 (147 ms), reached its minimum duration at 13 (143 ms) after which it had a significant increase to the adult duration of 159 ms. Within- and between subject variability both had a steady decrease up to 13 years of age and relatively stable values after that. Sentence duration also decreased with age and at the age of 14 years it is 45% shorter than at 7 years. At the age of 14 it is also shorter than at any other age and significantly different from the adults. Within-subject sentence durations stabilise at the age of 12 years.

A subset from the same data as above was used to further investigate vowel and consonant durations. Gerosa et al. (2006) measured duration of vowels, fricatives and stops in repetitions of five sentences by 35 children aged 5-17 (five children per year) and five adults. As in previous studies, the same effect of age on duration and within-speaker variability was observed. Vowels and stop consonants showed an almost linear decrease between seven and 13 years and relatively stable values after that, while fricatives exhibited a linear decrease throughout the observed ages. On average, the duration of vowels reduced by 41% between 7 and 17 years and duration of consonants by 25%. Moreover, greater variability was measured for vowels than for consonants although they both decreased with age.

Results from the above two studies give evidence that speech timing is not adult-like at the onset of adolescence but it continues well into the teenage years. During development the duration of segments, words and sentences is decreasing linearly with increasing age, reaching the lowest values between 13 and 17 years and converging to adult-like durations after that. Results additionally suggest a high between-speaker variability in temporal control as some children were able to produce adult-like durations already at the age of four years. However, the within-speaker variability on the duration measurements was the highest for the youngest children and it was decreasing with age until adult-like variability was observed around 12 or 13 years of age, suggesting consistency in articulation has been achieved.

Just as in adult speech, segments change their duration when they are part of a cluster in children's speech as well. Gilbert and Purves (1977) investigated onset consonants in English monosyllabic meaningful words with a (C)CVC structure in four groups of children and one adult group with five speakers per group. The children's group extended over the following ages: 5;0-5;6, 7;0-7;6, 9;0-9;6, and 11;0-11;6. Based on the measurements of segmental durations as singletons and in clusters (values given in Table 2.4) they observed that all age groups significantly shortened /s/ in /sl/ cluster compared to the singletons. The difference in the duration of /s/ as singleton and as part of a cluster ranged from 23.1 to 34.5 ms across groups and was not significant. In the case of /l/, the difference between the single and clustered realisation was significant only for the 5-year olds who lengthened clustered /l/ when compared to the singleton. Duration of /l/ was the same in both realisations for all the other age groups. The difference between the single and clustered /l/ was significantly different only between the youngest group and adults but it did show a linear increase with age. Speakers in the two youngest groups had on average clustered /l/ longer than singleton (5-year-olds by 21 ms, 7-year-olds by 7 ms) while all the other age groups had clustered /l/ shorter than singletons (9-year-olds by 0.2 ms, 11-year-olds by 1.4 ms and adults by 11.7 ms). Such results suggest that adult-like differences between single and clustered /s/ are present already in 5-year-olds, but the differences between the /l/ in the two conditions stabilise sometime between five and half and seven years. In accordance to other studies investigating temporal properties of children's speech, decrease in variability with age was observed for both single and clustered realisations of segments. The decrease is not linear but the youngest children had the highest variability and adults the lowest. The authors also conclude that durations of both single and clustered segments suggest a boundary between the youngest two groups of children and the older two groups and adults, and that all these results taken together imply a different timing control of segment duration between children up to seven years old and adults.

group	/l/	/l/ in /sl/	/s/	/s/ in /sl/	/s/ + /l/ in /sl/
5;0-5;6	94.8 (41.7)	115.8 (35.1)	194.5 (42.1)	160.0 (36.0)	275.8
7;0-7;6	87.0 (21.4)	94.0 (17.6)	196.3 (37.8)	173.2 (27.5)	267.2
9;0-9;6	77.4 (22.0)	77.2 (23.0)	168.7 (29.1)	142.8 (36.5)	220
11;0-11;6	69.7 (24.4)	68.3 (19.8)	166.1 (31.0)	132.8 (26.5)	201.1
adults	84.0 (19.35)	72.3 (14.1)	167.0 (25.7)	134.7 (22.3)	207

Table 2.4: Mean durations (and S.D.; both in ms) of segments by age group and onset structure as reported by Gilbert and Purves (1977)

Data presented by Gilbert and Purves (1977) does not only allow observing changes in segment durations when they are part of a cluster as compared to a singleton, but also changes in syllable onset durations when the number of onset segments increases. Summing the durations of /s/ and /l/ as part of word-initial /sl/ clusters gives us the duration of the entire syllable onset (Table 2.4). Although it is not possible to investigate the variability of cluster duration or statistically significant differences between the age groups, it can still be observed that the duration of the entire onset cluster decreases with age and that at least the youngest two groups exhibit much longer durations than the others. Decrease of the duration of consonant clusters with age was also found in a study by Cheng et al. (2007a) (details described in 2.3.2) which included speakers between six and 17 years of age.

Duration of simple and complex syllable onsets was especially explored in a study by Karlsson (2004). Eight Swedish speaking female children participated in the study and were recorded monthly between 1;7 and 3;1 years of age uttering simple onset syllables [lo:r], [no:r], [lo:r], and complex [sno:r] and [slo:r]. Both age of speaker and onset structure were shown to have a significant effect on the onset duration. Syllable onset duration increased with age and the mean onset duration for all realisations of all speakers showed that complex onsets had longer duration than simple onsets, with /sl/ being the longest. Out of the three simple onsets, /l/ was the shortest, followed by /s/ and /n/. Unfortunately, the author was only interested in the question whether there exist measurable differences between the duration of simple and complex syllable onsets in young children and reported the results for all speakers and ages together, with no information about the effect of the type of syllable onset on the onset duration. Following this, Karlsson (2006) investigated children's production of simple and complex onsets by expanding the age range from 1;2 to 4;2 years. He observed that the youngest children produced complex onsets with the shortest durations and the smallest variability of all the investigated ages. Both durations and variability are increasing until the longest durations are produced sometime between 2;6 and 3 years of age. After this both parameters are decreasing with age. Durations and variability of the single onsets stay more uniform over the investigated ages, and are lower than in the complex onsets.

2.2.2.3 Summary

Based on the studies presented above it can be concluded that children indeed have inferior speech timing control to adults. They produce longer segment durations and exhibit greater variability on a variety of speech units. All these parameters mature with age and become adult-like around the mid-teenage years. However, some aspects of speech timing mature earlier than others; e.g., duration is adult-like before variability. They additionally showed that although consonant clusters present a more difficult task for the maturing articulatory system, even very young children are able to apply the correct durations in order to successfully differentiate between simple and complex syllable onsets. Importantly, development is also not uniform across all children as some speakers achieve mature performance earlier than others. Group results of most of the reported studies have to be taken as a general guideline and not as set time points when examining maturation of temporal properties of children's speech.

Compared to typically developing children, speakers with CAS would be expected to show greater variability of segment and syllable durations, poorer adaptation of segment durations to the cluster environment, and inconsistent adaptation of syllable duration to the increased number of segments. Their segment and syllable durations would, however, be similar to those of typically developing children because of the high inconsistency masking any consistent differences between these speakers.

2.3 Tongue movements in speech of adults and typically developing children

A better understanding of the relation between duration and articulation can be obtained by direct observation of articulatory movements during speech. The following section presents some findings about tongue movements in adult speech and in the speech of typically developing children.

A number of human organs are included in speech production but this section will focus on tongue movements, because tongue was the articulator investigated in the research reported in this thesis.

The tongue is one of the most interesting parts of the human body. It is composed of a large number of muscles of different type and innervated with at least 6500 motoneurons per side (Stone et al., 2004) resulting in a high degree of freedom in its movement. The tongue can move up, down, left, right, can be rolled, bunched or flattened, with the sides moving uniformly or not. Anatomically it is a hydrostatic object which changes shape but preserves volume. A change in one part of the tongue has to result in a change in another part. The tongue has a crucial role in speech production and has always presented an interest to researchers. However, due to its position in the mouth, it has been a challenge to observe what exactly is going on during speech. For a long time, the acoustic signal was the only source of information about tongue position, with F1 being related to tongue height, and F2 to front/back position in the oral cavity. Access to a richer source of information became possible with the development of several techniques that enable direct observation of the tongue during speech. They primarily differ by the type of information they provide. X-ray and MRI, for example, enable observing not only the tongue but all the surrounding structures as well. Ultrasound images only the tongue surface, EMA traces pellets attached to the tongue, and EPG records time and location of tongue-palate contact. In the following section the focus is mainly on midsagittal tongue movements since that was the object of study in this thesis.

2.3.1 Adults

Studies investigating tongue movements in adults can be roughly separated into those exploring possible tongue shapes, positions and resulting movement patterns during speech, those revealing the existence of independent tongue regions and their correlations, and those describing tongue involvement in the realisation of individual segments.

Due to a high degree of freedom in tongue movement it is very interesting to observe how many movement patterns are actually present in speech. Interestingly, the study by Stone and Lundberg (1996) showed that there are only four possible midsagittal tongue shapes and three possible tongue-palate contact patterns. Ultrasound and EPG data of one 26 year old female speaker was recorded. She produced sustained vowels and consonants for ultrasound recording , and /pVp/ and /aC/ sequences for EPG recording. Ultrasound results showed that tongue shape can be described with only four distinctive categories: front raising, complete groove, back raising, and two-point displacement. The first three were observed in the production of vowels, and were additionally correlated to tongue position (e.g., front raising observed for front vowels). Consonants exhibited all four shapes. Front raising was observed in the production of /n/ and /ʃ/, complete groove in /θ/ and /s/, back raising in /ŋ/, and two-point displacement in /l/. The shapes were additionally formed in different position in the oral cavity. For example, front raising was observed in alveolar /n/ and in palatal /ʃ/. EPG data revealed even more differences between vowels and consonant. While only bilateral tongue-palate contact was observed for vowels, consonants exhibited also a cross-sectional pattern and the combination of the two. The authors concluded that vowels and consonants share the same tongue shapes, which are not dependent on the tongue position but rather reflect the aerodynamic needs of a segment and palatal morphology of a speaker. Vowels and consonants were more precisely differentiated by the tongue-palate contact pattern. This observation was confirmed in an earlier study using X-ray tracing of five pellets attached to the tongue (Stone, 1990). Results showed a correlation between tongue shape and position for the targets without tongue-palate contact, and no correlation for those with the contact. Furthermore, a higher number of posterior tongue shapes was observed for the segments with anterior bracing than for those without it.

Similar to Stone and Lundberg (1996), Green and Wang (2003) also identified only four distinctive tongue shapes following the analysis of X-ray tracing of four pellets in 46 participants uttering VCV sequences. The tongue shapes were similar to those described in Stone and Lundberg (1996) but seemed to be more related to the tongue positions. The four observed shapes were: body elevation (/j/), dorsum elevation (velars), blade elevation with dorsum depression (alveolars, palatoalveolars and retroflex), and anterior-blade elevation with body depression (/l/).

Speakers do not only employ a small number of tongue positions but are also very consistent in their production as Johnson et al. (1993) showed in a vowel production study. Six adults participated and their tongue shapes were observed by X-ray tracking of four tongue pellets. Speakers formed the same tongue shape at the vowel midpoints in each repetition (one to six repetition per speakers and vowel target).

In addition to observing the shape and movement of entire tongue, it is also important to observe individual parts of the tongue. The tongue does not function as a uniform object and its parts can move individually from each other. This was addressed in a study by Stone (1990) who used X-ray to trace movement of five tongue pellets, presenting five parts of the tongue, in different vowels and consonants. Results showed that the amount of tongue expansion and contraction were indeed not the same for all parts of the tongue. For example, in the case of /l/, the dorsal part of the tongue was the most expanded and the middle part the most compressed. /s/ had an even more compressed middle part and a largely expanded anterior one. The study also showed that, on average, the tongue was more compressed for vowels than for consonants.

Because speech is not a sequence of individual segments but rather a smooth blend of them, it is important to understand how tongue movements change in connected speech. This task was tackled by Iskarous (2005) who investigated tongue movement patterns in segment transitions. Tongue contours of 600 segment-to-segment transitions were extracted from an X-ray database. The patterns were evaluated by measuring the midsagittal distance function defined as the distance between the midsagittal tongue contour to palate and posterior pharyngeal wall. Results surprisingly revealed that 86% of all transitions showed only two basic patterns of tongue movements: pivot (Figure 2.1) and arch (shown in Figure 2.1 and Figure 2.2, respectively). The remaining 14% could not be assigned into these two categories but seemed to have an underlying pivot shape. Figure 2.1 and Figure 2.2 show tongue contours taken from /k/ to /a/ transition (Figure 2.1) and from /æ/ to /d/ transition (Figure 2.2). Red broken line contour is the first in the transition and the blue contour the last. In the case of /ka/ the tongue moves downwards through the transition, while in the case of /æd/ it moves upwards. The black curve presents the outline of the top and back of the oral cavity. Pivot pattern is defined by the stationary point in midsagittal distance function and not by the static point on the tongue. This means that in a transition between two segments there is a point in the oral cavity space that is constantly occupied by a part of the tongue. In contrast, the arch appears when the midsagittal distance function changes in one part of the tongue and the change becomes less and less until a region

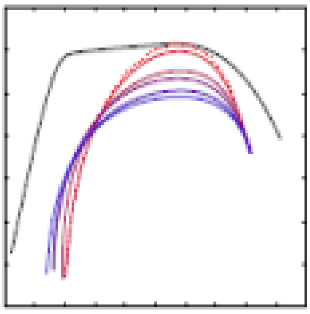


Figure 2.1: Example of a pivot pattern in /ka/ transition. Figure from Iskarous (2005).

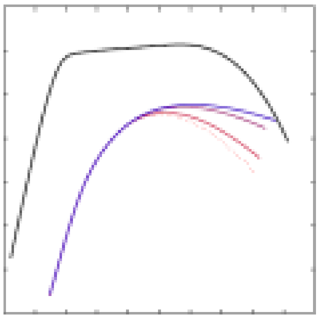


Figure 2.2: Example of an arch pattern in /æd/ transition. Figure from Iskarous (2005).

of no change is observed. Neither pivot nor arch pattern occur at the same part of the oral cavity, but depend on the place of constriction for the two adjacent segments. The author observed that the pivot pattern forms in transitions between segments produced by different or not tightly coupled parts of the tongue. It results from the release of the constriction of the first segment temporally overlapping the formation of the constriction of the second segment, with the constrictions being in different spatial locations. The arch, on the other hand, results in transition between segments with the same or tightly coupled parts of the tongue. In this case, the constrictions of the two segments are not only overlapped but also in the same location.

Changes in tongue movements in connected speech were explored also by Kuehn and Moll (1976) who demonstrated a connection between timing and velocity of speech movement, and proposed that articulatory movements may be primarily time-controlled. Ten male and seven female speakers participated in this study and produced repetitions of CV and VCVC sequences and a DDK task. Productions were filmed with X-ray and movements of various points on the articulators were measured. The main findings of this study were a high correlation between velocity and displacement of the articulators and very similar transition times between the articulators. Comparing measurements across subjects revealed a relatively constant transition time and highly variable velocity and displacement measures. Investigating measures of individual segments showed greater tongue velocities for /p/ and /l/ than for /s/. The authors conclude that the velocity of movement from C to V (in CV) depends on the amount of movement between the segments for all investigated consonants except /l/. Velocity for /l/-V transition was much higher than predicted from the amount of movement. The proposed explanation is that vowels are more dominant than consonants in CV transitions, resulting in tongue movements being more influenced by the demand to reach the appropriate tongue position for the vowel. /l/ is more vowel-like than other consonants and its gestures blend more with the following vowel. Slower velocity of /s/ is a result of more precise articulation than the one needed for stops.

2.3.1.1 Conclusions

The above studies agree that although the tongue can move in many different ways, as a uniform object or as a composition of individual parts, a very small number of shapes and movement patterns is observed in speech production. These shapes and movement patterns are additionally very stable in repetitions. Such functioning is not surprising and could be expected considering how complex articulation is from the point of view of motor planning and motor control. The speech production system is more likely to adapt and consistently execute the smallest possible amount of commands that enable producing intelligible output.

Analysing tongue movement patterns of speakers with CAS is likely to reveal two important results. First, they would produce qualitatively and quantitatively different and less smooth movement patterns than adults and more different shape types. Second, their patterns and shapes would be significantly less consistent in repetitions. Both these observations follow directly from the impairments of planning articulatory movements which characterise CAS, and can be most reliably uncovered through articulatory analysis.

2.3.2 Typically developing children

Because of the ongoing speech development process, tongue movements of children can be expected to be different from tongue movements of adults. The most common and least invasive way to investigate tongue movement patterns and shapes is to estimate them from the F1 and F2 frequencies of the acoustic signal. Typically, children have higher F1 and F2 frequency values and greater variability than adults but become adult-like with age, reflecting improvements in articulatory control (Kent, 1976; Nittrouer, 1993; Smith and Kenney, 1998). Variability is decreased to the adult level some time between 11 and 14 years of age, while formant frequencies reach adult values slightly later, between 13 and 15 years. Changes in F1 and F2 frequencies do not, however, occur linearly or at the same time for all children (Smith and Kenney, 1998). Although these results support the notion of a slowly maturing speech production system, they have to be taken with a level of caution. In general, children produce high fundamental frequencies, and as shown by Lindblom (1972, cited in Kent 1976, p.434) the higher the fundamental frequency of a speaker, the greater the chance for an error when measuring formant frequencies from a spectrogram. Because of this, it seems better to use articulatory than acoustic methods when assessing tongue movements in children's speech.

However, the application of articulatory methods is not very straightforward. They raise a high concern about the level of invasion. Because articulatory measurements allow direct observation of an articulator, they are always more invasive than a simple audio recording. The level of invasion stretches from the least invasive such as sticking of reflecting points to the face all the way to gluing coils to the tongue. Some of these procedures also involve a longer preparation phase or limit the participant's body movement during the recording and are thus less applicable to children. For these reasons articulatory analysis has not been as widely used to investigate speech of children as it has been to investigate speech of adults. Tongue movements have been investigated with EPG, ultrasound and EMA which are all harmless methods but do present different levels of invasion. EPG demands that the participant wears an artificial palate, ultrasound recording employs a probe and head stabilisation device, and EMA requires gluing coils to the tongue and face. This section gives an overview of the few studies investigating tongue movements in children's speech.

Ostry et al. (1984) used ultrasound to record 11 children, aged 3;3 to 11;6 years. The children produced disyllabic /kaka/ and /gaga/ sequences with stress on the second syllables. The authors measured the distance between the tongue dorsum and the transducer and described articulation with the relationship between distance of displacement, peak velocity and duration. Older children, aged 6;4 to 11;6 revealed an effect of voicing on displacement and peak velocity. Both parameters were smaller for /k/ than for /g/ targets which, according to the authors, has been reported for adults as well. Younger children (3;3 to 5;0), on the other hand, did not show the same effect. The age of participants had no effect on marking stress as all children, like adults, produced greater duration and dorsum displacement in stressed than in unstressed vowels. The average dorsum displacement in stressed vowels was 7.4 mm compared to 4.1 mm in unstressed vowels. They additionally report that one speaker had an average dorsum displacement, over a repetition of /kaka/, of 8 mm which is about two thirds of the movement observed in adults. This study also showed that even younger children already showed an adult-like consistent relationship between the distance of dorsum displacement and peak velocity, and no relation between the duration of the whole target and dorsum displacement. The authors concluded that children show the same kinematic relationships as adults and that through development their motor system does not change but only undergoes refinement. This would suggest that although young children achieve different values on articulatory and temporal measures, they are able to adapt their articulations in an adult-like manner.

Even more information about tongue movement was revealed in one of the biggest articulatory studies of children's speech (Cheng et al., 2007b,a,c). They used EPG to observe tongue-palate contact and EMA to observe tongue-jaw coordination in children and adults. Both studies included 48 participants divided into four age groups, 6-7 years, 8-11 years, 12-17 years and adults, with six males and six females in each group.

In the EPG study (Cheng et al., 2007b,a) the speakers produced repetitions of short monosyllabic words differing in the onset consonants and the authors investigated tongue-to-palate contact of /s/, /t/, /l/, /k/ both as singletons and as part of clusters /kl/ and /st/. The measurements showed a clear effect of maturation. Amount of tongue-to-palate contact was reduced with age, place of articulation of anterior consonants /t/, /s/ and /l/ was shifted forward, and the placement of tongue-to-palate contact in /k/ became more consistent (Cheng et al., 2007b). The children in the youngest group were additionally significantly more variable in the contact between tongue body and palate. An interesting observation of this study was also that a relatively substantial percentage of participants did not produce typical patterns for the production of investigated segments. This was especially evident in the case of /l/ where 17 % of the 6-7 year-olds, 8 % of the 8-11 year-old, 50 % of the 12-17 year-olds and 25 % of adults did not produce the typical EPG tongue-to-palate patterns. The authors suggested that speakers may have a greater range of tongue-to-palate placements available for the production of perceptually correct output then previously believed. With respect to tongue movement, the study showed no effect of age on the duration of the approach or release phase of tongue-to-palate contact, but a significant effect of age on the duration of closure phase (Cheng et al., 2007a). Within-speaker variability on any of these measures was not affected by age. The authors additionally observed a negative effect of age on the the overlap between a velar and alveolar gesture in /kl/ cluster, demonstrating a maturation of ability to separately control individual parts of the tongue.

Tongue motor control was further investigated by the same authors in an EMA study with the focus on the development of coordination between tongue and jaw during production of /t/ and /k/ embedded in a sentence (Cheng et al., 2007c). Analysis of the displacement and velocity of markers attached to the tongue and jaw showed a significant affect of age. They noted that in adults the tongue does not coordinate with the jaw as a uniform object. A clear separation between tongue tip (in /t/) and tongue body (in /k/) was observed. While tongue tip and jaw formed temporal coupling in their movement, tongue body and jaw remained independent but had a stable relation between their movements. Children, on the other hand, showed significantly less coupling between tongue tip and jaw and higher variation of coordination between tongue body and jaw. Both of these parameters underwent maturation until they became adult-like sometime between 8 and 11 years of age. However, similar to the acoustic and articulatory studies which included teenagers, some refinement of tongue-jaw coordination was reported to continue until late adolescence.

Variability in children's production can be additionally viewed in a study by Goozee et al. (2007) who recorded EMA data of four children, aged between 9;6 and 11 years. Children had two coils attached to their tongue and analysis was performed on word-initial /t, s, k/. EMA data revealed a relation between tongue tip and body movement during approach and release phase of the three consonants. On the majority of attempts, tongue tip and body moved in the same direction during the two phases. However, one child produced /t/ and /s/ with tongue tip and body moving in opposite directions both in approach and release phase, a second child did not move tongue body or moved it downwards in the approach phase of /s/, and a third child performed similarly in the approach phase of /k/. Although all of these children produced perceptually correct speech sounds, their tongue movements differed, supporting the observation that maturation processes are not yet completed at the onset of adolescence.

2.3.2.1 Summary

Tongue movements exhibit a significant effect of age on motor control. Younger children are less able to perform well timed and coordinated tongue movements. They are not able to differentiate between different sections of the tongue but move the whole tongue together when achieving articulatory goals (Cheng et al., 2007b). With age they increase their motor control and are able to separately control different parts of the tongue but different children reach maturity at different times (Cheng et al., 2007a; Goozee et al., 2007).

Similar problems in tongue movement control were observed in typically developing children, would be expected in speakers with CAS as well. However, they would probably show even greater variability by producing different types of movement patterns and shapes in repetitions.

2.4 Ultrasound imaging

In speech science, ultrasound has become a popular tool for studying tongue movements. Currently, it is the most non-invasive, safe, quick and low-cost technique used to obtain articulatory data on the tongue. Ultrasound depicts the surface of the tongue from a midsagittal or coronal view which enables the extraction of the tongue contour from one or several frames, visualisation of tongue movements, comparing tongue positions and measuring the amount of tongue movement between frames, duration analysis, and 3D reconstruction. Additionally, with the use of software packages, such as Articulate Assistant Advanced (AAA) by Articulate Instruments Ltd., ultrasound images can be time-aligned with the audio signal, recorded at the same time, which makes it possible to directly compare the tongue image and the acoustic properties of the speech.

This section first presents some general characteristics about ultrasound imaging before giving more detail about the application in the area of tongue imaging during speech and about the ultrasound system used at Queen Margaret University.

In general ultrasound enables the visualisation of tissue and internal organs and is medically most commonly used as a diagnostic tool, creating brightness-mode (B-mode) images. The B-mode ultrasound scanner emits ultrasound pulses into the body, receives the echoes and creates an image in which the strength of the reflection point is represented by the brightness of the image. The brighter the image of an anatomical structure, the stronger the ultrasound pulse reflected from that structure (Hangiandreou, 2003; Whittingham, 2003).

The ultrasound scanner consists of the probe, internal computer, screen and keyboard for manipulating the settings. A probe, which is both the transducer and the receiver, contains an array of piezoelectric crystals (most commonly there are 128) that transform electricity into ultrasound waves and vice-versa (Whittingham, 2003; Hangiandreou, 2003; Stone, 2005). The crystals emit high frequency pulses, ranging between 2 and 100 MHz (Blackwell, 2001). The frequency is directly related to the wavelength of an ultrasound wave which defines the resolution of imaged anatomical structures. As cited by Martin and Ramnarine (2003), the emitted frequencies of 2, 5, 10 and 15 MHz have a corresponding wavelength of 0.77, 0.31, 0.15 and 0.1 mm in soft tissue. The higher the frequency, the better the scanning resolution and the greater the ability to image small structures.

However, crystals do not emit the ultrasound pulses at the same time but in a successive order from one side of the probe to the other. The first crystal emits a new pulse only once the last crystal has received the echo from its pulse. Once the entire sweep is completed, an image is created (Hangiandreou, 2003). The speed at which the sweep of the entire array is completed defines the rate of the frames per second the scanner can create.

Another important probe feature is the shape of the area containing the crystals. Probes can have either a linear or curvilinear arrays of crystals (Whittingham, 2003). Linear probes enable a rectangular field of view both at the surface (at probe-skin contact) and deeper in the tissue. Curvilinear probes, on the other hand, produce a wide field of view at the surface but even wider deeper in the body. A special type of curvilinear probe is the strongly convex curvilinear probe which displays a narrow field of view on the surface and a wider view deeper in the body. The selection of the probe is according to the imaging target and its position in the body.

Figure 2.3 shows a strongly convex curvilinear probe held under the chin and emitting ultrasound waves upwards towards the tongue and receiving ultrasound waves reflected from the tongue surface.

Once the ultrasound pulses are emitted, they travel onwards through the medium (e.g. tissue, air, water) until they reach a boundary between mediums with different density. At such a boundary, the waves are reflected back and detected by the probe. The greater the density difference between mediums, the stronger the echo. The amount of the reflected waves additionally depends on the angle at which the emitted pulse hits the boundary. The best echo is achieved when the pulse and the boundary are orthogonal to each other.

In an ideal situation a pulse is emitted by the probe, travels onwards through the tissue in a straight line, orthogonally hits the tissue-to-air boundary, is reflected back and detected by the probe. Based on the time at which the pulse was sent, the time at which

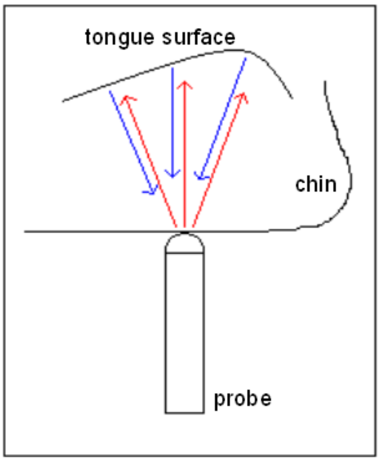


Figure 2.3: The probe emitting and receiving ultrasound waves in tongue imaging. Ultrasound waves are emitted upwards (red arrows) by the probe held under the chin. After reaching the boundary between tongue surface and air or bone above it, they are reflected downwards (blue arrows) and detected by the probe.

it was detected by the probe and the speed of sound, the ultrasound machine calculates the distance the signal has travelled and the point of reflection. However, the tissue is usually not that uniform and the differences inside it present small-scale reflecting targets which cause an ultrasound wave (in its original form or the echo) to become scattered, usually over a large range of angles (Martin and Ramnarine, 2003). The ultrasound waves scattered by small targets are much weaker than ultrasound waves reflected by large targets. Because of this, not all of the scattered echoes are detected by the probe. However, in case they are detected, they can create phantom images. Phantom images occurring in tongue imaging are illustrated in Figure 2.4 below. They appear because the distance of travel of the signal is always calculated as if it has travelled in a straight line. By doing that, it is possible to calculate a point of reflection which does not exist. Phantom images and "lost echoes' can also result from refraction. When an ultrasound wave travelling trough one medium encounters and passes through another medium, its direction of propagation changes. Consequently, its echoes cannot reach the probe or the apparent points of reflection are not correct.

Furthermore, as the ultrasound pulse travels through the medium its energy is transformed into heat (absorption) and its intensity decreases over the distance (attenuation). Absorption is dependent on the frequency of the ultrasound (Martin and Ramnarine, 2003). The higher the frequency the quicker the signal is absorbed and therefore the shallower is the scanning depth. For this reason, most of the medical ultrasound imaging is done with frequencies between 2 and 15 MHz (Martin and Ramnarine, 2003; Blackwell, 2001) to achieve a compromise between the scanning depth and the resolution of the targets.

After all the echoes are returned and the points of reflection are calculated, the shape of the reflection area is composed and shown on the screen as a 2D grey-scale image with the brightest points resulting from the strongest echo. The number of captured frames depends on the scan rate of the scanner. A 30 Hz scanner thus captures 30 frames per second, creating a frame every 33 ms. The ultrasound's output can be either a single image taken at a specific time point or a video image recorded over time.

However, because of the characteristics of video format, video recordings of ultrasound images cause some misalignments of scanned images. According to Wrench and Scobbie (2006) the majority of ultrasound scanners operate on the continually updated B-mode image memory. This means that each time a new sweep starts, the data from the old one is still stored in the ultrasound memory. When such images are stored into video format, either NTSC comprised of 486 lines of image or PAL comprised of 576 lines, the lines are filled at slightly different time points and thus originate from different ultrasound images. Remember that a new sweep continues overwriting the previous and, at some point, starts being overwritten by the following sweep. As explained by the authors, if the ultrasound frame rate is 1.5 times faster than the video frame rate, the resulting video image is composed of three ultrasound sweeps. The left side of the image is created by interlacing the current and previous ultrasound image, the centre of the image by the current ultrasound image only, and the right side by the current and following image. If NTSC (29.97 fps) format is chosen, the resulting video image spans 33 ms, and in the case of PAL (25 fps) 40 ms. Transforming ultrasound images into video does not present a problem when imaging static or very slowly moving objects and for that reason ultrasound is still described as a real-time imaging method. However, it does influence ultrasound imaging of fast moving structures such as the tongue during speech. This is further discussed below in section 2.4.1, describing characteristics of ultrasound imaging of the tongue.

2.4.1 Ultrasound imaging of the tongue

Just as with any other application of ultrasound, imaging of the tongue requires several adaptations regarding the position and the surrounding environment of the imaged object. However, unlike other applications of this methodology where the imaged object is relatively still, it has to additionally deal with fast movement of the tongue.

For the purpose of tongue imaging, the probe is placed underneath the chin and the waves propagate through the tongue until they reach its surface which creates a boundary either with the palate (e.g., back of the tongue in /k/) or the air (e.g., in /a/). The tongue surface is represented by the lower edge of the white curve in Figure 2.4. The tongue tip is on the right side of the ultrasound image and the tongue back is on the left.

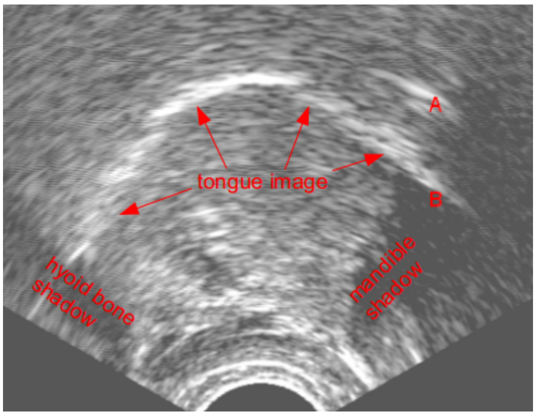


Figure 2.4: Ultrasound scan of the midsagittal tongue contour showing hyoid bone and mandible shadows, and two phantom images. One phantom image (A) is above the imaged tongue surface and one (B) inside the mandible shadow.

In the case of the tongue-air boundary, 99 % of waves are reflected back because of the difference in density between the tongue and the air. For example, more waves are reflected at tissue-to-air and tissue-to bone boundaries than from tissue-to-tissue or tissue-to-water (Stone, 2005). Just as in any other ultrasound imaging, tongue imaging is affected by ultrasound refraction and scattering. Refraction is most typically caused by interdigitated muscles and scattering by fat (Stone, 2005) and can result in the creation of phantom images. Figure 2.4 shows two phantom images. Phantom image (A) was created above the imaged tongue surface. Because of the tongue surface tissue-to-air boundary at which the majority of ultrasound waves is reflected, it is very unlikely that another structure could be imaged above the tongue. Ultrasound waves also cannot travel through bones and for that reason structures above the bones cannot be imaged and any image created (B) is a phantom as well.

In fact, bones create so-called acoustical shadows. In midsagittal tongue imaging, two acoustical shadows are typically created, one resulting from the mandible and the other from the hyoid bone (Figure 2.4). They can reduce the amount of tongue surface visible in the scans, either by obstructing the tip or the root of the tongue.

The wide variety of media (different types of tissues, air and bone) present in tongue imaging is also a cause of the greatest limitation of the technique. When the tongue tip is raised, an air pocket is created below it and the ultrasound waves are echoed already at this tissue-to-air boundary and cannot reach the tongue tip. For this reason it is not always possible to say whether the end of the tongue surface image is also the end of the tongue tip or the tongue tip is actually not captured.

Another important factor in tongue imaging is stabilisation of the transducer and head positioning. The transducer has to be kept in contact with the chin all the time during the recording and orientated in the same direction, so that the resulting images present the tongue from the same view. This is usually achieved with a specially designed transducer and head support system or a headset which fixes the probe and restricts head movements but does not limit speech movements. A good transducer and head stabilisation also prevents the occurrence of artefacts such as double edges (e.g., one curve resulting from the lateral side of the tongue and another from the tongue's groove) and discontinuities on the resulting image (Stone, 2005).

The quality of an ultrasound image can be additionally reduced by the type and structure of the tissue of the speaker's chin. With age, people tend to accumulate more fat in the chin. Unfortunately, fat has relatively high density and so fat in the chin scatters the ultrasound waves before they reach the tongue surface. Tongue imaging is made difficult in such cases.

As mentioned previously, one more important issue influencing ultrasound imaging of the tongue is storage of recordings in video format. Because the tongue is a relatively fast moving object, the resulting video images (composed of three ultrasound frames) display double images, discontinuities and shape distortions. During 33 ms spanning one NTSC video image, the tongue tip can move 10 mm and the tongue body 7mm (Wrench and Scobbie, 2006). However, as recommended by Wrench and Scobbie (2006), there are a number of steps that can improve the accuracy of measurements. They suggest making sure that older images do not influence the current ones by setting the system's persistence to zero. To achieve a better frame rate NTSC format is preferred. The frame rate of the ultrasound system should be set by selecting the minimum depth (around 8 cm) and minimum field of view that covers the target area (90 - 120°).

2.4.2 Ultrasound system at Queen Margaret University

The ultrasound system available at Queen Margaret University and used in this study consists of the ultrasound machine DP6600 by Mindray, endocavity probe model 65EC10EA (5, 6.5 and 8 MHz; 128 pyzoelectrical crystals in the head), PC, Multichannel System and Probe Stabilisation Headset by Articulate Instruments Ltd., prompt display screen, clip-on microphone. The ultrasound machine and PC are located adjacent to the recording room in order to exclude the noise produced. A diagram of the system is represented in Figure 2.5 and Figure 2.6 shows the headset fitted to the head and with a probe fixed under the chin.

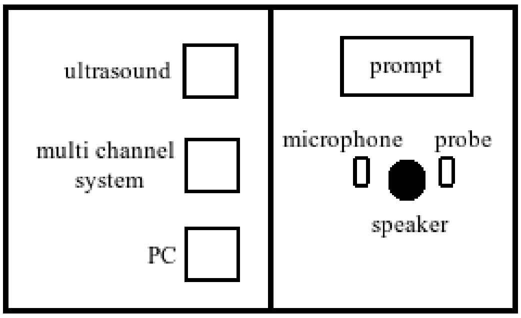


Figure 2.5: Ultrasound system set-up in Queen Margaret Ultrasound recording studio.

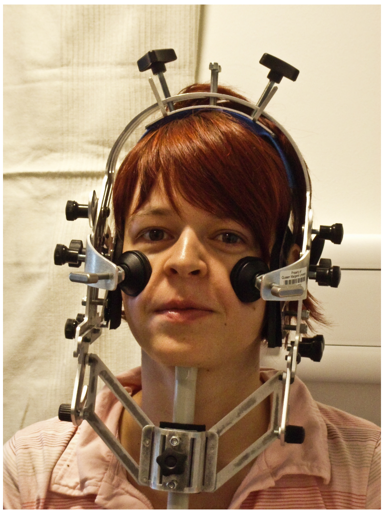


Figure 2.6: Headset fitted on the head and with fixed ultrasound probe under the chin.

Recordings are done with the AAA software which enables the ultrasound scans to be transformed into NTSC video format and to be synchronised with the audio recording. Unfortunately, this synchronisation was not yet perfect during the data collection for this thesis. In fact, the AAA software has undergone a high number of changes in this time and the data was recorded and analysed on several versions available between May 2007 and October 2009. The synchronisation was, however, improved by adjusting for the possible misalignments. It was observed in a previous experiment by Vazquez Alvarez and Hewlett (2007) that the ultrasound image showing tongue position at a specific time in the audio signal can appear up to 40 ms before or 40 ms after the selected time point. A slight improvement of synchronisation was made in the software prior to analysing data for this thesis. It meant that an ultrasound image could be only delayed relative to the audio signal. The exact time difference between the signals was not known, but as reported earlier by Zharkova and Hewlett (2009), it could be expected that it did not significantly influence measurements of tongue movements over syllables longer than 100 ms.

The AAA software does not only allow recording but analysis of recordings as well. Both audio and ultrasound signal can be inspected and/or annotated in order to select desired ultrasound frames. The next step in ultrasound data analysis is to fit the tongue contour to the lower edge of the bright curve representing tongue surface. This is done semi-automatically by first specifying the area inside which the software looks for the brightest lower edge. Once the AAA finds such edge, a contour is fitted to it. This fitting is, however, not perfect and all the frames have to be inspected and the fitted contours manually corrected. An example of a fitted tongue contour is given in Figure 2.7. The red curve on the image represents the tongue surface contour, and the green and grey curves specify the boundaries inside which AAA looks for the lower brighter edge and fits the tongue contour. Once all the frames are fitted, the tongue contours can be either used directly in the AAA to inspect their shape and compare different tongue contours or they can be extracted as a series of (x,y) points in the Cartesian or polar coordinate system. These points are extracted at the crossings of the tongue surface contour and the lines of the fan. The fan has in total 84 vertical lines but only seven are represented in Figure 2.7 (fan colour transitions from blue to white). The number of extracted (x,y) points depends on the length of the tongue contour; the longer the tongue contour, the higher the number of extracted points. Having the outline of the tongue surface or the set of points representing it, allows comparing the shape of the tongue in different frames or measuring the distance between tongue position from different frames.

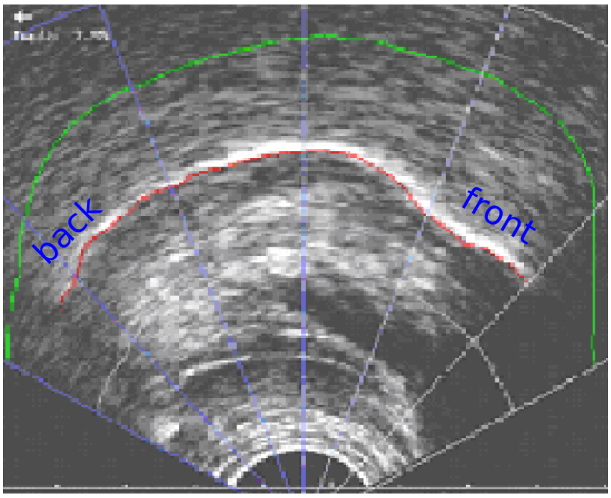


Figure 2.7: Example of fitted tongue contour (red) to the ultrasound image of the tongue. The green and grey curves specify the boundaries inside of which AAA searches for the image of tongue surface.

2.5 Summary of Chapter 2

Three main conclusions important for the study described in this thesis can be made from the presented review of earlier work. First of all, it clearly brings out the complexity of CAS. Because the unknown, underlying impairment causes a disruption at the level of planning and organising movement sequences, the resulting speech is affected in more than one way and CAS is for that reason more difficult to recognise and define. Previous work on CAS has shown that it cannot be diagnosed only on the basis of the most notable, and relatively simple to recognise, markers such as the type and number of segmental errors or deviant stress patterns. It cannot even be easily recognised on the basis of acoustic information. The exploitation of articulatory methods, however, seems to be more promising because of their power to directly reflect articulatory movements during speech, and because they enable combining acoustic and articulatory data. Even more, recent data showed that understanding of CAS would additionally benefit from addressing speech perception abilities of affected speakers. In conclusion, the more low-level and diverse are the methods used in investigations of CAS, the most likely it is that critical characteristics will be revealed and later used in the diagnosis and treatment procedures.

The second conclusion of this review is that adults and typically developing children clearly differ in their ability to control temporal and tongue movement properties in speech. Because one of these groups possesses mature and the other one immature, but typical, speech, they seem to be good candidates for control groups to the speakers with CAS who have immature speech due to the impairment. The relationship between the three groups can reveal new characteristics about CAS as compared to adult and typically developing speech. Additionally, the relationship between the two control groups can serve as an indicator of the validity of the applied methods, compensating for the relatively small amount of research focused on CAS. If the methods used show expected results for the two control groups, they most likely show correct results for speakers with CAS as well.

The third conclusion is that ultrasound imaging is an appropriate method for researching speech characteristics of motor speech impairments like CAS. Because it enables direct observation of tongue movements and synchronised observation of changes in the articulatory and acoustic domains, it provides a richer set of data that can potentially reveal new information about CAS.

The next section gives details about the main motivation for the research presented in this thesis, research questions addressed and the hypotheses predicting the relationship between the group of speakers with CAS and the two control groups, and the relationship between the control groups.

2.6 Main questions and hypotheses addressed in this thesis

The main motivation for the study reported in this thesis was the idea that if CAS is a truly motor speech disorder, it might be more valid to directly investigate articulation. Observing tongue movements seemed the logical choice because of the tongue's high level of involvement in articulation. Additionally, temporal characteristics of speech were analysed. Addressing articulatory and acoustic properties of speech together, provides richer information and can help in revealing the main characteristics of CAS.

Reviewing previous findings on speech characteristics in CAS, ultrasound as a research tool, and tongue movements in adults and typically developing children, led to three main questions, each with associated hypotheses, as follows:

Q1: Do speakers with CAS have impaired timing? Are their duration measurements different from those of adults or typically developing (TD) children?

- Effects of number of syllable onset segments on syllable duration
  - H1: Adults and TD children show an effect of syllable onset structure on syllable duration by significantly increasing syllable duration with the addition of an onset segment. Speakers with CAS do not show the same effect.
- Differences in segment duration according to whether segment is the only element of syllable onset or is part of an onset cluster
  - H2: Adults and TD children consistently adjust segment duration when the segment is part of an onset cluster as compared to the segment being the only element of syllable onset. Durations of segments change significantly in the following way: /p/ and /s/ lengthen when followed by /l/; /p/ and /l/ shorten when preceded by /s/; /s/ shortens when preceding /p/; and /l/ shortens after /p/. Speakers with CAS do not show the same adjustments.
- Duration of syllables: comparison of the three groups of speakers
  - H3: Adults have significantly shorter syllable durations than TD children. Speakers with CAS do not differ systematically from either of the control groups.

- Variability of syllable duration: comparison of the three groups of speakers
  - H4: Adults have significantly lower variability of syllable durations than TD children and speakers with CAS. Speakers with CAS are additionally significantly more variable than TD children.

Q2: Do speakers with CAS have impaired tongue movements? Are their tongue movements different from those of adults and typically developing children?

- Effects of the type of syllable onset segments on the amount of tongue movement over the syllable
  - H5: Adults and TD children show a consistent effect of the type of syllable onset segments on the amount of tongue movement, in that the addition of a lingual segment significantly increases the amount of tongue movement over the syllable. Speakers with CAS do not show the same effect and can either increase, decrease or not change the amount of tongue movement.
- The amount of tongue movement over the syllable: comparison of the three groups of speakers
  - H6: Adults have significantly smaller amount of tongue movement over the syllable than TD children. Speakers with CAS do not differ systematically from either of the control groups.
- Variability of the amount of tongue movement: comparison of the three groups of speakers
  - H7: Adults have significantly lower variability of the amount of tongue movement than TD children and speakers with CAS. Speakers with CAS are additionally significantly more variable than TD children.
- Patterns of tongue movement: comparison of the three groups of speakers
  - H8: Individually, adults and TD children show similar tongue movement patterns when articulating similar syllables but adult speakers have more consistent realisations across repetitions than children. Speakers with CAS show different tongue movement patterns than adults and TD children and are at the same time more varied across repetitions.

Q3: Do speakers with CAS have impaired rate of movement? Is their rate of movement different from that of adults and typically developing children?

- Effects of syllable onset structure on the rate of tongue movement over the syllable
  - H9: Adults and TD children show a consistent effect of the type of syllable onset and number of onset segments on the rate of tongue movement over the syllable. The addition of a lingual segment does not significantly change the rate of tongue movement, while the addition of a non-lingual segment significantly decreases it. Speakers with CAS do not show the same effect. The addition of either a lingual or non-lingual segment can increase, decrease or not change the rate of tongue movement.
- The rate of tongue movement over the syllable: comparison of the three groups of speakers
  - H10: Adults have significantly higher rate of tongue movement than TD children. Speakers with CAS do not differ systematically from either of the control groups.
- Variability of the rate of tongue movement over the syllable: comparison of the three groups of speakers
  - H11: Adults have significantly lower variability of the rate of tongue movement than TD children and speakers with CAS. Speakers with CAS are additionally significantly more variable than TD children.

Chapter 3

Methodology

Chapter 3 presents the participants, speech material, recording procedure, measurements and analysis procedures used to test the above hypotheses and to answer questions about motor and temporal components of speech in CAS. Results of the investigated speech characteristics in the three speaker groups, adults, TD children and speakers with CAS, and of individual speakers with CAS are presented in Chapter 4. and discussed in Chapter 5.

3.1 Participants

The original aim was to record 10 adults, 10 typically developing children and 10 teenagers with CAS. The goal was achieved for the adult and child groups, but not for the CAS group. Only three speakers with CAS were included in the study. All speakers were native speakers of English coming from the Edinburgh area, and had no speech, language, hearing or cognitive impairments, apart from CAS in the CAS group. Their details are given in Table 3.1. Prior to data collection, the study was evaluated by Queen Margaret University Ethics Committee and the National Research Ethics Service (documents related to the recruitment of participants can be seen in Appendix I. The former approved the inclusion of adults and typically developing children, and the latter of CAS speakers who were all recruited through a NHS facility.

3.1.1 Speakers with CAS

Teenagers with CAS were recruited through the department of Child Life and Health at the Royal Hospital for Sick Children, Edinburgh, with the kind help of Prof Anne

participant	group	gender	age
AD1	adult	f	29
AD2	adult	f	25
AD3	adult	f	21
AD4	adult	f	22
AD5	adult	f	22
AD6	adult	f	28
AD7	adult	f	26
AD8	adult	f	20
AD9	adult	f	28
AD10	adult	f	23
TDC1	child	m	7;5
TDC2	child	f	8;7
TDC3	child	f	6;10
TDC4	child	f	6;5
TDC5	child	m	9;3
TDC6	child	m	6;7
TDC7	child	f	9;4
TDC8	child	m	9;2
TDC9	child	f	6;8
TDC10	child	m	8;11
CAS1	cas	m	13;11
CAS2	cas	m	18;10
CAS3	cas	m	18;10

Table 3.1: Information about the participants: speaker's code, group, gender, and age (given in years for adults and years;months for children and speakers with CAS).

O'Hare. She inspected medical files of potential candidates, selected those that satisfied the inclusion criteria and provided contact details. The inclusion criteria included the age of potential participants between 14 and 19 years, diagnosis of CAS since early childhood, presence of CAS speech characteristics regardless of speech therapy received and no other known language, hearing, cognitive or neurological disorders. All the referred clients had been diagnosed with CAS by speech and language therapists working at the department. The speech and language therapists all had a high level of experience working with different speech and language disorders, particularly with those of neurological origin. During the recruiting procedure I was very aware of the issues involved in the selection of speakers with CAS as this has presented problems and controversy since CAS was first described as an independent speech disorder. Lack of any clear set of speech characteristics and no verifiable neurogenic impairment leave the decision about the diagnosis of CAS mainly to the expertise of speech and language therapists and their ability to exclude other potential disorders. Due to confidentiality issues I was not allowed to review medical or speech and language therapy records of potential participants. For that reason, the best available option was to trust members of staff working at a well-established and experienced department, such as Child Life and Health, to select potential participants. I believe that by letting experienced speech and language therapists select and recommend their clients improves the certainty of correctly diagnosed CAS. Additionally, selecting older participants (14 - 19 years old, mean age 17;2) who have been diagnosed as children and had the diagnosis confirmed over the years, reduces the risk of misdiagnosis. In total, eight teenagers with CAS were selected by Prof Anne O'Hare and given information about the study but only three speakers decided to take part. All of them were males and two of them, CAS2 and CAS3, were identical twins.

At the time of ultrasound recording, participants with CAS were asked to complete the Oral and Speech Motor Control Protocol by Robbins and Klee (1987). The protocol assesses 24 vocal tract structures and their relationships, 56 vocal tract functions, both in the speech and non-speech domain, maximum phonation time and repetitions of monosyllabic and polysyllabic sequences. The protocol was originally tested on children aged between 2;6 and 6;11 but it can be used to asses speech and non-speech motor control in older speakers as well. All three speakers with CAS scored the maximum number of points on the structural part of the protocol, which excluded any structural deviations in their oral cavity. They were equally successful on the functional part of the protocol assessing speech motor control. All three speakers produced perceptually correct segments and words. However, they showed more problems on the part assessing non-speech motor control functions, although they all completed the tasks. Speakers CAS2 and CAS3 had some difficulties elevating their tongue to the alveolar ridge. The elevation was not completely smooth and they showed some uncertainty about the direction of the movement. The same was additionally observed in the anterior to posterior palate sweep of CAS3. Results of the maximum phonation time and DDK rates are presented in Table 3.2 together with the norms for Robin and Klee's (1987) oldest TDC group (6;6-6;11).

participant	maximum phonation time (s)	/pa/ (number/s)	/ta/ (number/s)	/ka/ (number/s)	/pataka/ (number/s)	"patticake" (number/s)
CAS1	9.56	3.7	2.3	2.3	2	3
CAS2	3.54	3.3	3.3	3.7	3	3
CAS3	1.26	3	3	3.7	4	4
TDC 6;6-6;11	11.47 (SD 3.02)	5.73 (SD 0.43)	5.37 (SD 0.72)	4.85 (SD 0.71)	1.72 (SD 0.19)	1.64 (SD 0.26)

Table 3.2: Results of the three speakers with CAS for maximum phonation time (s), and for the repetitions of monosyllabic and polysyllabic sequences (number of repetitions per second). The table additionally includes normative data (mean and SD) for the 6;6-6;11 TDC group as reported by Robins and Klee (1987).

As can be seen, only CAS1 achieved comparable maximum phonation time to the normative TDC group, while the other two speakers with CAS performed below the norms for the 6;6-6;11 group. Speakers CAS2 and CAS3 additionally showed variation of pitch and loudness in the maximum phonation time task. All three speakers also produced less repetitions of individual syllables on the DDK task than the TDC group. CAS2 had additional problems with producing incomplete closure in /t/ in the fast repetitions of /ta/. Incomplete closure was not noticed in his production of isolated segments and words or in spontaneous speech. Interestingly, however, all three speakers with CAS produced more repetitions of polysyllabic sequences, both real and non-words, than the oldest group reported by Robbins and Klee (1987).

In addition to the above tasks all speakers also produced some spontaneous speech in the form of a dialogue and a short narration. They all participated in the conversation about what they expect from being a participant and they all described one of their hobbies. Speaker CAS1 was less talkative (mainly due to being shy as pointed out by his parent) than the other two participants. All three were intelligible and, as assessed perceptually, none of them showed any problems with controlling pitch or loudness, none of them made any noticeable segmental errors or changes in voice quality.

The protocol itself is not reliable enough to allow any clear conclusions about the motor control characteristics of speakers with CAS but it still gives an insight into some of their abilities. None of them had any problems on the speech task involving individual speech sounds and words. This is not very surprising, since all of them have been receiving speech therapy and did not make any segmental errors in spontaneous speech. What was more surprising was that they all performed poorly on the maximum phonation and monosyllabic repetition task but not on the polysyllabic one. Their slower rate and shorter phonation times could be attributed to their speech impairment which disrupts speech motor planning, timing control and prevents speakers from producing quick sequences or sustaining a particular oral tract configuration. Poorer performance on DDK is one of the most common speech characteristics used to diagnose CAS. However, it was expected that they would have more problems also with the repetition of the polysyllabic sequence which demands quick planning of different motor goals. Another possible explanation could also be that participants were shy about "singing" or repeating nonsense words. This was observed particularly for the phonation task but although they were asked to repeat the task three times and encouraged to sustain the vowel as long as they could, the phonation time did not improve.

3.1.2 Adult and child speakers

In order to investigate how timing and motor control differ in speakers with CAS, two control groups were selected: adult speakers (between 20 and 30 years of age) and typically developing children (between 6 and 9 years of age; mean 7;11) without any speech, language, hearing or cognitive disorders. The reason for including younger children was mainly to enable comparison of tongue movements of speakers affected by CAS, and of those affected by typical speech development processes, and to explore in what ways both these groups of speakers differ from standard adult speakers. Suitability of participants was verified by asking them, and their parents in the case of minors, about any known violations of the inclusion criteria. Adults and typically developing children were both recruited through Queen Margaret University student and staff mailing lists. All the adult participants were females. This was not intentional and it reflects a high percentage of female students and staff at Queen Margaret University.

Out of 10 children, five were girls and five boys.

When recording these two groups of participants, some limitations of ultrasound as an investigation technique became obvious. The main problem for adult speakers was that the tongues of some participants did not image well by ultrasound, probably because of the amount and structure of chin tissues (as mentioned in section 2.4). When inspecting the recorded ultrasound images it was observed that for some speakers the tongue image was not created in all the frames. For example, a speaker might be discarded because in their realisation of "pay", three out of 12 ultrasound frames had a very unclear or missing tongue image. Such frames make it impossible to trace the tongue outline for the entire syllable and were discarded. For this reason, 17 adults had to be recorded to obtain 10 recordings in which the tongue surface could be traced in all the scanned frames of the target syllables.

More participants had to be recorded in the children's group as well. In total, 15 children took part to obtain 10 good quality recordings. The main problem with children was not so much the structure of chin tissue but a long probe handle. The probe used in this study had a very long handle and the handle stuck down from under the speaker's chin. The handle was too long for some children to enable them to sit comfortably. When they started moving around, the probe handle got caught in their legs or hands and was moved. The probe angle relative to the chin was therefore changed which resulted in changed scanning angle. When inspecting the recorded ultrasound images it became obvious that the scanning angle was changed so much that the resulting image did not represent the same part of the mouth. Such recording would not allow comparisons between the tongue contours and had to be discarded. Unfortunately the probe with the long handle was also the one with the smallest probe head which could be fixed most appropriately, and comfortably for the children, under the chin. Additionally, although the recording took only up to 15 minutes, this turned out to be too long for some children as they could not sit still.

3.2 Speech material

Several criteria were employed when selecting speech material. First, the words had to differ in the complexity of the syllable onset with the same consonants appearing as singletons and as part of a consonantal cluster. Second, the words had to be monosyllabic to reduce the effect of coarticulation. Third, all the words had to end on the same vowel. In this way the speakers had the same articulatory target at the end of each of the words, allowing comparing tongue positions across different words. Finally, the words had to be real English words. Following this, six monosyllabic English words were selected: “pay”, “say”, “lay”, “play”, “slay” and “splay”. They differed in the number and type of onset segments, which included one non-lingual (/p/) and two lingual (/s/, /l/) consonants. In order to provide the same environment for all six target words, they were uttered in a carrier phrase “a ____ today”. In this way it was also made sure that speech material was not too big to allow comfortable recording session for the participants.

3.3 Recording procedure

Before the recording it was explained to the participants how the ultrasound machine works and what would they have to do during the recording. Participants also had time to become familiar with the probe and headset which was particularly welcomed by the children.

The first step in the recording procedure was fitting the headset, including the probe (see section 2.4.2), which took between 10 and 15 minutes, depending on the amount of adjustments needed and the participant’s cooperation. Since this was the first study in which the headset was used on children, some concern had been raised over its size, weight, and appearance. Fortunately, none of the children refused to take part after seeing the headset or having if fitted and the adaptation to the smaller heads of children was achieved by adding more side padding. When asked, all children said that they could feel the weight of the headset but that it was not uncomfortable for the time of recording, which did not exceed 15 minutes.

The acoustic signal was captured through a small clip-on microphone, fastened to a headset to keep it at a more or less constant distance from the speaker’s mouth throughout the recording, no matter what head movements occurred. This was particularly necessary when recording children, who moved more than adults.

The internal frame rate of the ultrasound machine used in this study was set to 30 fps. However, due to the transformation of scanned ultrasound images into the NTSC video format, the resulting video frame rate was 29.97 fps. The frequency of the emitted ultrasound pulses was set to 5 MHz. This was previously chosen by the members of the ultrasound laboratory group at Queen Margaret University as it was the best compromise between resolution and depth of pulse penetration. The depth setting was assessed prior to the recording and the smallest one enabling viewing the entire tongue during speech was chosen (following recommendation by Wrench and Scobbie (2006)). For all AD and TDC participants it was set to 7.55 cm. All three CAS speakers were males and had bigger heads for which scanning depth had to be changed to 8.62 cm. The probe was fixed under the chin according to the instructions in Articulate Instruments Ltd. (2008) and in such a way that the centre of the image was the midpoint between the shadows caused by the hyoid and jaw bone and with a field of view spanned over 120°.

Once the headset was fitted, the participants were sited in front of a computer screen where prompt words were displayed together with the preceding /ə/ (Figure 3.1 shows example of screen with the prompt). Prompt words were displayed in a randomised order during each recording session. The order was defined by assigning each prompt with a number, using a random number generator to produce a sequence of the numbers and then ordering the prompts according to it. This was done for every recording session. Participants were instructed to utter all the displayed prompts with the addition of the word "today" at the end. Every participant made five or six repetitions of each target. Because some of the children did not feel comfortable reading, they were given cues, either just the first consonant or the entire syllable, before uttering the words. Two such children (TDC3 and TDC9) were included in the analysed data set.

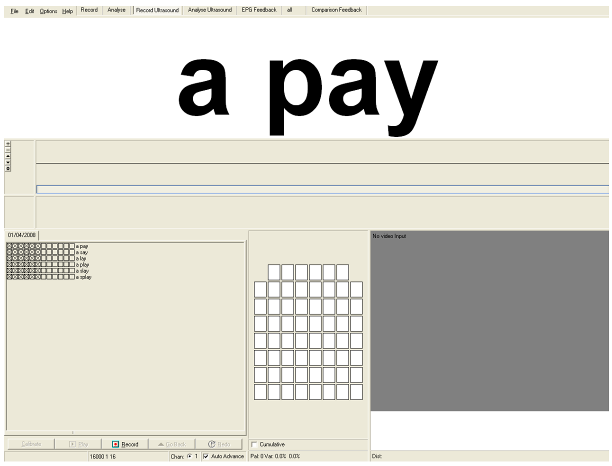


Figure 3.1: Prompt screen as viewed by the participants.

3.4 Data analysis

After data collection, the recordings were inspected and a decision was made whether the quality of the obtained images was good enough for analysis. Data was discarded if the tongue image was very faint or absent on more than two consecutive frames of the target syllables or if the angle of the probe changed during the recording. Selected speakers were singled out and five repetitions of each of the six targets per speaker were selected and included in the analysis. The only exception was speaker AD3 who contributed only four repetitions of “play” because of a failed recording of one attempt.

3.4.1 Duration

The first parameter measured on the recorded data was duration of target syllables and their segments. The syllables were first annotated in the AAA and then, together with the accompanying acoustic signal, exported into Praat (Boersma and Weenink, 2008) and segmented. Praat allows better spectrogram displays, more precise annotations and automatic extraction of different measures. Both syllable annotation and segmentation were based on the acoustic waveform, spectrograms and perceptual evaluation. The segments were marked according to the following criteria:

- /p/: Beginning marked by a drop in waveform amplitude and disappearance of the spectrogram characteristics of the previous segment. End marked by the appearance of a pitch period after the burst and of formant structure of the following segment.
- /s/: Beginning and end marked by continuous spectrogram energy in higher frequency region.
- /l/: Beginning marked by a lowering of F1 and F2, end by the characteristics of the following segment.
- /e/[1]: Beginning marked by the start of a periodical waveform and the presence of clear formant structure on the spectrogram. Because the vowel was always followed by /t/, its end was marked by a drop in waveform amplitude and disappearance of the formants.

After all the data was annotated and segmented, syllable durations were measured in the AAA while segment durations were extracted from Praat.

[1] /e/ is a standard symbol representing Scottish English version of Standard English /eɪ/ in “pay”

3.4.2 Amount of tongue movement

In order to measure the amount of tongue movement, tongue contours have to be fitted to all the relevant frames of the annotated syllables. As described earlier, contour fitting is done semi-automatically by the AAA and corrected by hand. Because video images are composed of more than one ultrasound image (2.4), which results in overlayed tongue surfaces, fitting the tongue contour was not always straightforward. In such cases, the image was first compared to the previous and the following frame to help in deciding the exact shape of the tongue surface. If that did not provide enough information, the image was de-interlaced. This allowed comparing tongue movement from the "problematic" frame in two separate frames. Once the decision was made about the shape of tongue surface, the contour was fitted to the original, interlaced, frame.

Another useful feature of the AAA is that it can create a new frame by interpolating between two original frames. This was applied to obtain frames presenting the most likely tongue position at the beginning and end time points of a syllable. Because the frame rate of the used ultrasound system was 30 fps, it was very unlikely that a frame was created exactly at those two time points. In order to create new frames, tongue contours were fitted also to the frame preceding the first frame of the annotation, and to the frame following the last frame of the annotated region. A new frame could then be created at the selected time point by interpolating between "real" frames from either side of that time point.

Having the tongue surface traced in all the frames inside the annotated region, allowed assessing the potential misalignment of the acoustic and the ultrasound signals by visually inspecting and comparing tongue movements over the same syllables. It was reported earlier that although there is a chance of up to 40 ms difference between the two signals recorded by AAA, the tongue movements over syllables longer than 100 ms are likely to be unaffected (Zharkova and Hewlett, 2009). The inspection of data obtained here confirmed this, as it showed that individual speakers, especially adults, had similar midsagittal tongue contour patterns in the five repetitions (see figures in section 4.4.1).

All the fitted tongue contours were exported as a series of (x,y) points and the following analyses were based on them and executed in Matlab. The basis of measuring the amount of tongue movement was the average nearest neighbour distances (aNND) measured between every pair of consecutive tongue contours of the utterance (contour

1 and 2, 2 and 3, 3 and 4...) (Zharkova and Hewlett, 2009).

The first step in obtaining the amount of tongue movement measure was cutting pairs of consecutive contours to a similar length by finding the nearest neighbour distance (NND) between the end points of the contours and cutting at the points which were closer together. Figure 3.2 shows a pair of contours differing in length. First, the NND was measured from all four end points (1 and 3, A and C). It was found that 1-B NND is shorter than 1-A, and 2-C shorter than 3-C. Because of that, points 1 and B were selected as cutting points on the left end of the contours and points 2 and C on the right end. Following this, the contours of each pair were each represented by 100 equally spaced (x,y) points. This was achieved by interpolation. After that, aNND could be calculated. aNND is an average of all the NND measured between the points on the two contours of a pair of tongue contours . To obtain aNND, NND was measured from every point on the first contour of the pair to the second, and from every point on the second to the first. Figure 3.3 illustrates measuring NND from red tongue contour to the blue (red lines between the points on the contours), and from blue to the red (blue lines between the points on the contours). Because the contours had 100 points each, the result was 200 NND, averaged to give aNND. The amount of movement was then obtained by summing aNND of all pairs of consecutive tongue contours of the target.

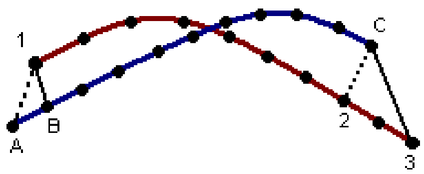


Figure 3.2: Cutting of a pair of tongue contours of different length to the same length.

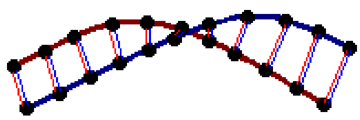


Figure 3.3: Measuring NND from the red tongue contour to the blue, and from the blue to the red.

The first attempt at measuring the amount of tongue movement followed the above description exactly, with distance including aNND between every consecutive pair of contours for each target. However, because different syllables included different numbers of ultrasound frames (the longer the duration of the target, the higher the number of scanned frames) it was believed that some kind of a normalisation was necessary, especially because with a greater number of frames, the chance of making an error when fitting the tongue contour also increases. To my knowledge, this study was the first ultrasound study of this kind in which a number of consecutive frames was included in the analysis. For this reason there was no known method of achieving some kind of data normalisation. After discussions with other members of the department working with ultrasound, it was decided that the best way was to make sure that all the syllables were represented by the same number of ultrasound frames. In the final analysis, only the frames from the following nine time points were included:

1. beginning of first consonant of the C1(C2)(C3)V1 syllable
2. five equally spaced points inside the onset to capture the tongue position of all consonants
3. boundary between onset and V1
4. mid V1
5. end of V1

In order to have the same number of ultrasound frames in all syllables, a Praat script was written to first insert the new time points into the audio signal. The audio and the new annotations were then imported back into the AAA where new frames were created at these time points by interpolating between the adjacent original frames resulting from the ultrasound imaging. Tongue contours were fitted to the new frames as well, which resulted in nine tongue contours per syllable target. Amount of tongue movement was still measured in the same way as described above, by taking consecutive pairs of the target syllable (eight pairs).

3.4.3 Rate of tongue movement

Rate of tongue movement was calculated by dividing amount of tongue movement over the entire syllable by syllable duration:

$$rate = \frac{amount}{duration}$$

3.4.4 Tongue movement patterns

To obtain a graphic representation of tongue movement patterns all the nine tongue contours of each syllable were plotted together to allow observation of the changes in tongue position between segments. Another way of graphically representing tongue movement patterns was achieved by plotting the highest point on each of the nine tongue contours of the syllable. These kinds of plots allowed inspection of how consistent tongue positioning was over repetitions of the same syllable and observation of the differences between the three groups of speakers. Because of the complexity of patterns and amount of data involved, tongue movement patterns are analysed only by describing these figures. No quantitative analysis was performed. Tongue movement patterns are shown individually for the three speakers with CAS, three AD, and three TDC. The three speakers from each of the non-clinical groups were selected because they had the greatest amount of tongue surface visible on the ultrasound image.

Tongue movements were additionally inspected by observing tongue contour patterns occurring in transitions between segments, as described by Iskarous (2005) (details are given in Section 2.3.1). All transitions between segments in the second repetition of the six syllables of each of the selected speakers were inspected and assigned either a pivot, arch or combined pivot/arch transition pattern.

3.5 Statistical analysis

In order to reveal within-speaker and between-speakers differences across syllables, descriptive statistics were calculated for each of the syllable types. All the measurements were first tested for whether they followed a normal distribution, by using Shapiro-Wilk's test of normality and by inspecting quartile-by-quartile plots. Both methods showed that the majority of measurements violate the assumption of normality and for that reason non-parametric techniques had to be applied. Descriptive statistics are thus represented by median and interquartile range (IQR) values. In order to establish if syllable-type influenced duration, amount of tongue movement and rate of tongue movement, two statistical procedures were applied: first, a Wilcoxon Signed-Rank Test and second, generalised linear mixed models (GLMM).

A Wilcoxon Signed-Rank Test was used for paired comparisons. Each syllable was compared to every other syllable, which resulted in up to 15 comparisons per measured parameter. The high number of comparisons called for the application of the

Bonferroni correction, which reduces the alpha level in order to reduce the possibility of wrongly rejecting the null hypothesis. The Bonferroni correction reduced the alpha level from .05 to .001 - 0.008. However, reducing the alpha level, in turn, increases the chance of a type II error, of not accepting the difference between the groups even when the difference is true (Field, 2009). The decision of how many tests have to be taken into account when correcting the alpha level remains an open question.

As discussed by Cabin and Mitchell (2000), the published studies do not always make it clear whether all tests were performed on the same data set or on subsets. Additionally, their review of articles published in Ecology and results of questionnaires sent to the editors of the journal revealed that there is also no clear consensus whether corrections of the alpha level should be applied to the separate data sets or the whole data presented in the article. They also concluded that such thinking can lead to conclusions that data has to be corrected for all the tests ever done or that a researcher has to keep correcting alpha values for all the tests performed in her lifetime. Nakagawa (2004) pointed out that a lack of consensus over when to use a Bonferroni correction results in researchers reporting only those tests that show significant results, even though if the total number of tests was reported and alpha levels corrected accordingly, the significance would be no longer observed. Because of this, suggestions have been made to stop applying Bonferroni corrections and rather find alternative ways of handling the increased chance of a type I error (Nakagawa, 2004) or to report all the data and results as they are and let the reader decide about the level of adjustments she thinks is necessary (Saville, 1990).

Additionally, when comparing syllable types between groups, the median value of each speaker's five repetitions was used, ensuring that the real number of speakers was represented in the analysis. Wilcoxon Signed-Rank test does not account for repetitions nested under speakers but treats each repetition as a separate speaker. If all five repetitions per speaker were included, analysis would be performed as on 50 instead of ten speakers (for adults and TD children groups).

Although the Wilcoxon Signed-Rank test is a well established non-parametrical statistical procedure to evaluate differences between groups, a large adjustment of the alpha level and inclusion of only median values of speakers' repetitions in the between-groups comparisons caused some doubt about it being the best choice. Examination of the data suggested differences between groups of speakers but the necessary alpha level failed to reflect them. Additionally, it was believed that including all repetitions into the analysis is more appropriate to better encompass by-speaker and by-syllable variability.

For these reasons, a recent addition to the field of speech science, GLMM (Baayen et al., 2008; Jaeger, 2008; Baayen, 2008), was applied to the data. Because GLMM are highly complex statistical procedures, a more detailed explanation is beyond the scope of this thesis. However, a brief overview of GLMM and its application, based mainly on the descriptions in Baayen et al. (2008) and Baayen (2008), is given below.

GLMM enable discovering factors that significantly influence a measured parameter (dependent variable) by modelling it as a linear combination of fixed and random factors. As argued by Baayen et al. (2008), language studies have to treat participants as random factors since the main goal of the study is to reveal how a certain phenomenon behaves in all the language users and not just in the participating subjects. Participants differ from each other on a wide variety of characteristics, from genetic, to social, educational, and environmental and including them in a statistical model as a random factor covers a bigger spectrum of possible personal characteristics. For the same reason, the speech material used has to be treated as a random factor. Again, most of the studies try to answer more general questions about language or speech characteristics and not just the differences and similarities of the language material. In general, both speakers and speech material could be selected randomly within the necessary inclusion criteria. It should be possible to use different speakers and different words to replicate a study. Because of defining speakers as a random factor, GLMM do not demand averaging of the repetitions prior to the analysis but can include the variability present in the repetitions. Similarly, they can handle missing data and unbalanced data sets. GLMM additionally do not ask for the assumptions of normality within groups or equal variance between groups and can be used on data that would normally have to be analysed with non-parametric statistical methods. Because GLMM do not directly test the differences between groups of data but analyse which factors have a significant influence on the data, no adjustments to the alpha level are needed. Moreover, GLMM produce estimates of the intercepts and slopes of fixed effects for an average subject and an average item while at the same time capturing individual differences by estimating intercepts and slopes of individual subjects and items. They also allow inclusion of all the possible factors that are expected to influence the dependent variable or identifying irrelevant factors by comparing models with different specifications.

In this study GLMM were applied by using the lmer function in the statistical package R (R Development Core Team, 2009). In all the model specifications both speaker and syllable were specified as random effects, syllable duration, segment duration, amount of tongue movement and rate of tongue movement as the dependant factors and the number of onset segment, the number of lingual or non-lingual onset segments and the type of onset segment (single segment or cluster) as the predictors.

Summaries for model objects fitted with lmer do not list alpha levels expressing statistical significance, but only standard error and t-statistics. In general, statistical significance at the 5% level can be estimated by calculating the 95% confidence interval. Reported standard error is multiplied by 1.96 and then first subtracted from the estimate for the intercept to obtain the lower interval boundary, and added to the estimate to obtain the upper interval boundary. If the 95% confidence interval includes zero, it is very likely that there is no relation between the predictor (e.g., the number of syllable onset segments) and the dependent factor (e.g., syllable duration). If zero is not present, it can be concluded with only 5% chance of an error that the predictor truly has an effect on the dependent variable. However, this method was reported to be over conservative for small data sizes. In such a situation it is more appropriate to apply Markov Chain Monte Carlo (MCMC) sampling from the posterior distribution of the parameters which allows extraction of alpha values. Because of the relatively small data size used in this thesis, the alpha values obtained through GLMM will be based on the posterior distributions (pMCMC) and reported together with upper and lower higher posterior density intervals (HPD).

Due to the complexity of the data both procedures, traditional Wilcoxon Signed-Rank Test and GLMM, were applied and are reported in the results section. It was believed that combined they can better characterise the data and address the proposed hypotheses.

Chapter 4

Results

This chapter will first look at the measures of syllable and segment durations, amount of tongue movement and rate of tongue movement within the AD, the TDC and the CAS groups, before comparing them between the groups. It is important to keep in mind that the AD and TDC groups both consisted of 10 speakers and the CAS group of three as this difference may influence group results. The next section of the chapter will present the patterns and consistency of tongue movements in the three speakers with CAS and selected three AD and three TDC speakers. Finally, the last section looks at the results of individual speakers with CAS.

Information about statistical tests testing for the significance of the results is given in this chapter only in the form of obtained *p*-values. Full test results are given in Appendix II.

The first step in the measures analysis was to choose the appropriate statistical measures and procedures by testing whether the data followed a normal distribution. Table 4.1 shows the results of a Shapiro-Wilk test of normality for the distribution of syllable duration, amount of movement and rate of movement, respectively. All the measures are broken down by speaker group and syllable. The tests revealed that although most of the syllable durations and amounts of movement were distributed normally ($p > 0.05$), this was not the case for all measures and especially not for the rate of tongue movement. Existence of data that was not normally distributed demanded the use of the non-parametric Wilcoxon Signed-Rank and Mann-Whitney tests. As explained previously in Chapter 2, results were additionally investigated by applying GLMM.

Speakers	AD			TDC			CAS		
Syllable	dur *p*-values	aom *p*-values	rom *p*-values	dur *p*-values	aom *p*-values	rom *p*-values	dur *p*-values	aom *p*-values	rom *p*-values
pay	0.293	0.419	0.021 *	0.004 *	0.098	0.001 *	0.023 *	0.021 *	0.857
say	0.193	0.005 *	0.672	0.769	0.712	0.539	0.275	0.448	0.469
lay	0.044 *	0.008 *	0.000 *	0.041 *	0.463	0.008 *	0.071	0.525	0.362
play	0.014 *	0.006 *	0.000 *	0.169	0.649	0.027 *	0.048 *	0.104	0.312
slay	0.278	0.466	0.000 *	0.329	0.720	0.013 *	0.007 *	0.139	0.040 *
splay	0.368	0.253	0.000 *	0.096	0.415	0.004 *	0.802	0.045 *	0.288

Table 4.1: *p*-values obtained by the Shapiro-Wilk test of normality evaluating distribution of syllable duration (dur), amount of tongue movement (aom) and rate of tongue movement (rom) by syllable type and speaker group. *p*-values $>$ 0.05 indicate normal distribution and are marked by (*).

4.1 Group results: Duration

4.1.1 Syllable duration

4.1.1.1 Within speaker groups

Syllable duration was first inspected by observing the effect of the number of onset segments on syllable duration within speaker-groups and comparing the effect across groups. Durations of each of the six syllables in the three speaker-groups are shown as boxplots in Figure 4.1. Group median value is represented by a dark horizontal line, IQR by the height of the box, the range between the most and the least extreme values by whiskers extending from the box, and the outliers by circles. Median and IQR values are additionally shown in Table 4.2. Presenting the same data in figure and in table format contributes to uncovering the main characteristics of each group and relation between the groups.

Speakers	AD		TDC		CAS	
Syllable	Median	IQR	Median	IQR	Median	IQR
pay	321	110	424	156	420	130
say	332	124	457	149	414	115
lay	255	117	337	95	330	80
play	348	112	468	149	436	110
slay	378	121	563	155	503	125
splay	422	116	570	300	535	103

Table 4.2: Median and IQR values of syllable duration (ms) by syllable type and speaker group.

Figure 4.1 and Table 4.2 illustrate that all three speaker groups showed at least some increase in syllable duration with the addition of a syllable onset segment for all six possible pairs: pay-play, say-slay, lay-play, lay-slay, play-splay, and slay-splay. However, not all of these differences proved to be significant when applying the Wilcoxon Signed-Rank test and Bonferroni correction of the significance level. Because of the multiple paired tests performed on this data, the standard significance level of 0.05 was reduced to $p < 0.008$. p-values resulting from the Wilcoxon Signed-Rank test are shown in Table 4.3 and all those that proved significant are marked by (*).

As can be seen, AD significantly increased syllable duration with the addition of a segment in all tested pairs, except "slay" - "splay". The TDC group showed slightly less consistency, and additionally made no difference in the durations of "pay" and

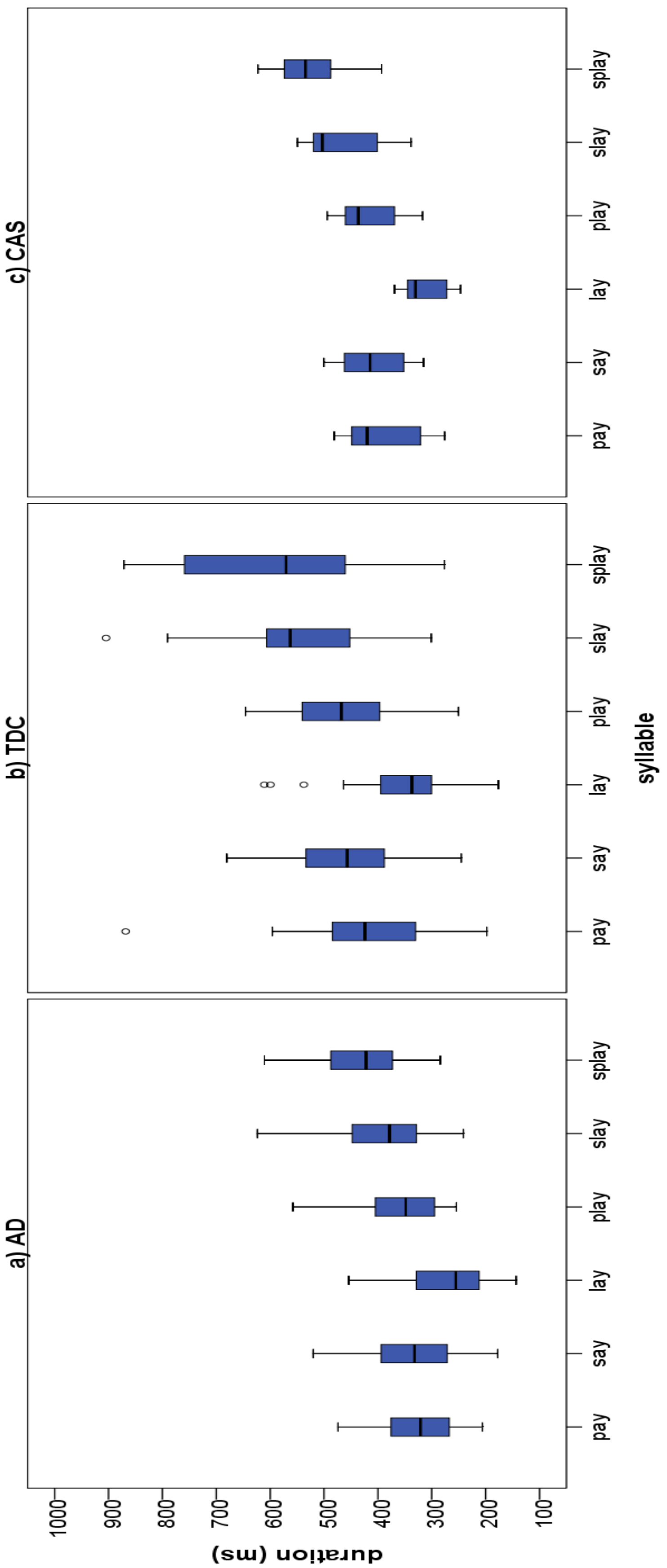


Figure 4.1: Syllable durations (ms) in each of the speaker groups: (a) AD, (b) TDC, (c) CAS

	Speakers	AD	TDC	CAS
		p-values	p-values	p-values
WSR	pay-play	0.0025 *	0.0110	0.1090
	say-slay	0.0025 *	0.0045 *	0.1090
	lay-play	0.0025 *	0.0025 *	0.1090
	lay-slay	0.0025 *	0.0025 *	0.1090
	play-splay	0.0025 *	0.0025 *	0.1090
	slay-splay	0.0140	0.0110	0.1090
GLMM	Number of onset segments	0.0014 *	0.0032 *	0.0001 *

Table 4.3: p-values obtained by a Wilcoxon Signed-Rank (WSR) test evaluating differences in syllable duration between syllable pairs and by GLMM modelling of the effect of the number of syllable onset segments on syllable durations in each of the speaker groups. p-values < 0.008 in the case of WSR and < 0.05 in the case of GLMM are marked by (*) and indicate significance.

”play”. CAS speakers showed no effect of the number of onset segments on syllable duration.

The effect of the number of syllable onset segments on the syllable duration was further assessed with GLMM. This procedure enabled the modelling of syllable duration as a function of the number of onset segments and with speakers and syllables as random factors. In contrast to the Wilcoxon Signed-Rank test, modelling showed that the number of onset segments had a significant effect (p-values < 0.05) on syllable durations in all three speaker groups (Table 4.3). All groups significantly increased duration with the addition of a segment to the syllable onset.

4.1.1.2 Between speaker groups

Another way of inspecting differences in syllable duration between speaker groups was to compare duration and within-group variability of duration of individual syllable types. Both measures can be observed in Figure 4.1 and Table 4.2 which clearly show that the AD group had the shortest duration for all syllable types and the TDC the longest. Variability measures did not show such a distinction between the groups. The CAS group was the least variable in the duration of “say”, “lay”, “play” and “splay”, and the AD group for the remaining two syllables. The highest variability of duration was observed in the TDC group for all the syllables except “lay” where the AD one produced more variable utterances. Overall, the CAS group showed the least variability on most of the targets and the TDC the most.

In addition to the measure of variability, minimum and maximum values measured in each of the groups were inspected as well. They are marked by the end-points of whiskers in Figure 4.1. As can be seen, speakers in the TDC group achieved higher minimum values than the AD group on “say”, “lay” and “slay”, but similar on the remaining three syllables. All their maximum values were higher than in the AD group. The CAS group achieved similar maximum values as the AD group on “pay” and “splay” and lower maximums on the other syllables. They also had higher minimum values on all syllables than the AD group. Similarly, their ranges of syllable durations stayed well inside the minimum-maximum range of the TDC group.

Differences between speaker groups were first tested by applying the Mann-Whitney test. The resulting *p*-values obtained by the test can be seen in Table 4.4, with those showing significance marked by (*). The alpha level was reduced to 0.003 to correct for multiple paired tests. As can be seen, none of the syllables had significantly different duration in a comparison of the three groups. Almost the same was observed for within-group variability of duration (Table 4.5). “lay”, “play” and “splay” were the only syllables with greater variability in the TDC group than in the AD. No other differences in variability of duration were observed across the three groups.

Speakers	AD vs. TDC		AD vs. CAS		TDC vs. CAS	
Syllable	MW *p*-values	GLMM *p*-values	MW *p*-values	GLMM *p*-values	MW *p*-values	GLMM *p*-values
pay	0.0170	0.0030 *	0.1760	0.0360 *	0.8660	0.5620
say	0.0095	0.0001 *	0.1760	0.0276 *	0.3980	0.2264
lay	0.0170	0.0008 *	0.3100	0.1060	0.4990	0.2570
play	0.0245	0.0002 *	0.2370	0.0636	0.6120	0.2938
slay	0.0065	0.0002 *	0.1760	0.0318 *	0.3100	0.1332
splay	0.0080	0.0001 *	0.0630	0.0212 *	0.4990	0.2482

Table 4.4: *p*-values obtained by a Mann-Whitney (MW) test evaluating differences in syllable durations between speaker groups and by GLMM modelling of the effect of speaker group on syllable durations for each of the syllable types. *p*-values < 0.003 in the case of MW and those < 0.05 in the case of GLMM are marked by (*) and indicate significance.

Different results were obtained when syllable durations were modelled as a func-

Speakers	AD vs. TDC		AD vs. CAS		TDC vs. CAS	
Syllable	MW *p*-values	GLMM *p*-values	MW *p*-values	GLMM *p*-values	MW *p*-values	GLMM *p*-values
pay	0.0050	0.0082 *	0.0880	0.5590	0.1550	0.2874
say	0.0870	0.0624	0.1550	0.5590	0.4330	0.4662
lay	0.0020 *	0.0052 *	0.4330	0.7324	0.0215	0.0904
play	0.0015 *	0.0016 *	0.1990	0.6576	0.0315	0.0686
slay	0.1130	0.1096	0.4330	0.8220	0.1990	0.1850
splay	0.000 *	0.0001 *	0.0140	0.0308 *	0.1550	0.0858

Table 4.5: *p*-values obtained by a Mann-Whitney (MW) test evaluating differences in the IQR of syllable durations between speaker groups and by GLMM modelling of the effect of speaker group on the IQR of syllable durations for each of the syllable types. *p*-values < 0.003 in the case of MW and those < 0.05 in the case of GLMM are marked by (*) and indicate significance.

tion of speaker group. As reported in Table 4.4, the TDC group produced all syllables significantly ($p < 0.05$) longer than the AD group. Speakers with CAS had longer durations than the AD group for four of the syllables, "pay", "say", "slay" and "splay", but not for "lay" and "play". Exploring the effect of speakers belonging to either the TDC or the CAS group revealed no significance. The TDC and the CAS speakers produced the same syllables with similar durations. In addition, the AD and the TDC groups differed the most on variability of duration. Table 4.5 shows that, as a group, TDC were significantly more variable on "pay" , "lay", "play", and "splay" than AD. The CAS group showed greater IQR than the AD one on "splay" only, and was as varied on all targets as the TDC group.

4.1.2 Segment duration

Figure 4.2 shows durations of /p/, /s/, and /l/ as singletons and as part of a cluster. Exact median and IQR values are given in Appendix II, Table 4.

Similarly to the measures of whole syllable durations, the AD speakers had the shortest segment durations of the three speaker groups. Neither of the other two groups had consistently longer segment durations across all targets. The TDC group had the longest durations of /s/ and /l/ in all contexts, and speakers with CAS of /p/. Some additional difference between the groups was observed in the adaptation of segment

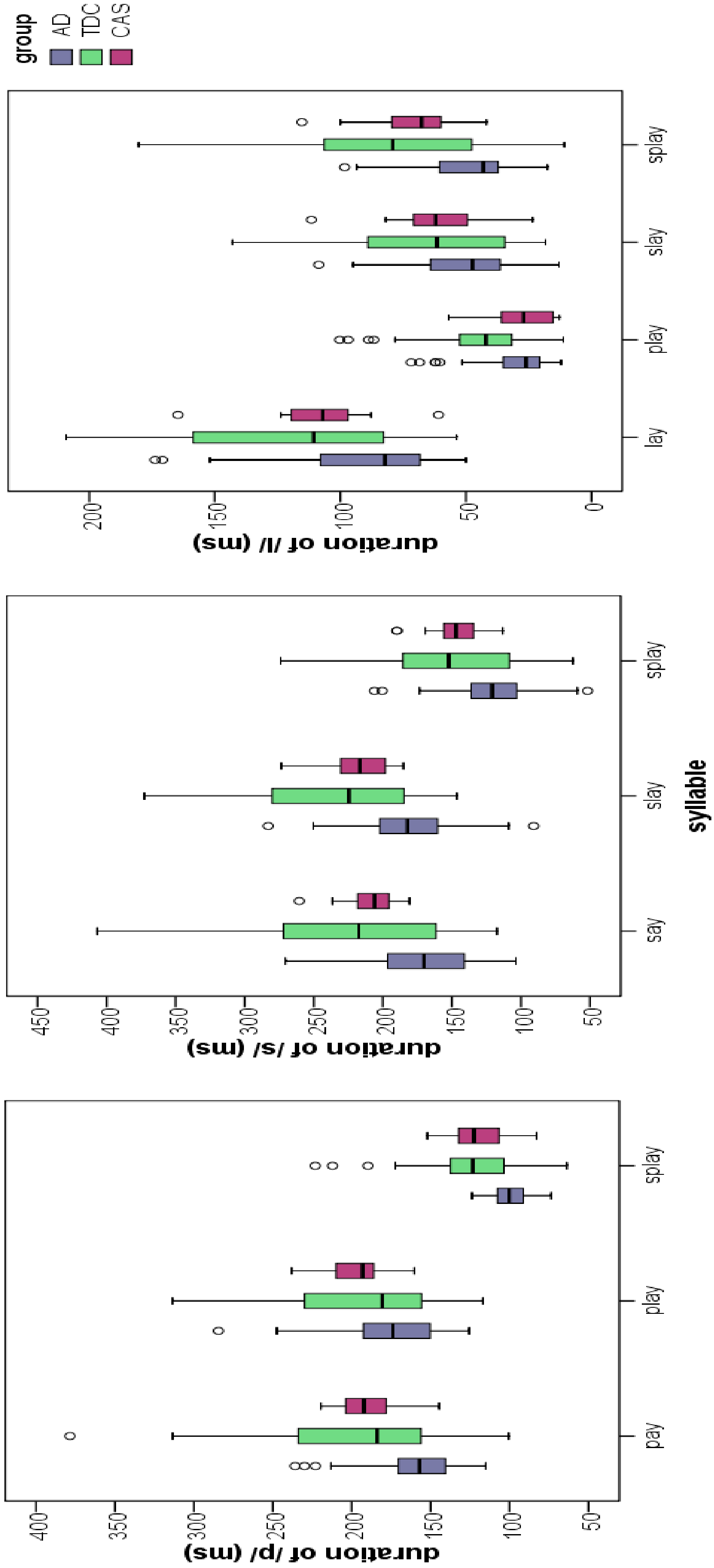


Figure 4.2: Durations (ms) of /p/ (left), /s/ (middle) and /l/ (right) as single- and clustered- syllable onset segments in each of the speaker groups. Note that the scale is different in each figure.

durations in different onsets. As can be seen, the three speaker groups changed the duration of some segments in the same way while showing differences on others. All groups increased /s/ when followed by /l/ in “slay” as compared to being the only onset segment (in “say”), and decreased it when part of a three segment syllable onset /spl/. They also all produced the longest /l/ when it was a single onset element and reduced it in clusters. /l/ was the shortest in “play” in all three groups. In the AD group /p/ had longer duration in “play” and shorter in “splay” than in “pay”. In contrast, /p/ in “play” was slightly shorter than in “pay” in the TDC and the CAS groups. Both of these groups also decreased duration of /p/ in “splay” as compared to “play”.

Statistical significance of the differences in segment durations in different syllable onsets was again tested with a non-parametric Wilcoxon Signed-Rank test. The alpha level had to be corrected and the differences were significant only when $p < 0.005$. Obtained p-values can be seen in Table 4.6 and show that none of the tested pairs differed significantly in segment duration in any of the speaker groups.

Contrasting results were obtained when data was modelled with GLMM. p-values obtained by the model (Table 4.6) revealed no effect of the onset type (singleton or cluster) on the duration of /s/ in any of the speaker groups and no effect on the duration of /p/ in the AD and the CAS groups. The TDC group, on the other hand, produced a different duration of /p/ when /p/ was a singleton than when it was part of a cluster. Similarly, duration of /l/ was significantly affected by the syllable onset structure in all three groups.

4.2 Group results: Amount of tongue movement

4.2.1 Within speaker groups

Characteristics of the amount of tongue movement over the syllable were first evaluated by observing the effect of the addition of a lingual segment to the syllable onset in the three speaker groups. As evident in Figure 4.3 and in Table 4.7, all three speaker groups showed an increase in the amount of tongue movement with the addition of a lingual segment to the syllable onset. However, they also showed an increase when non-lingual /p/ was added (“lay” - “play”, “slay” - “splay”). Overall, the tongue had a shorter amount of movement over all syllables with single-segment onsets than over syllables with onset clusters and the greatest amount over “splay” for all three speaker groups.

		Speakers	AD	TDC	CAS
	Segment		*p*-values	*p*-values	*p*-values
WSR	/p/	pay-play	0.005	0.508	0.109
		pay-splay	0.005	0.005	0.109
		play-splay	0.005	0.005	0.109
	/s/	say-slay	0.047	0.139	0.109
		say-splay	0.005	0.005	0.109
		slay-splay	0.005	0.005	0.109
	/l/	lay-play	0.005	0.005	0.109
		lay-slay	0.005	0.007	0.109
		lay-splay	0.005	0.009	0.109
		play-splay	0.007	0.005	0.109
		slay-splay	0.799	0.074	0.109
GLMM	Type of syllable onset	/p/	0.2620	0.0341 *	0.2522
		/s/	0.3506	0.4680	0.2308
		/l/	0.0132 *	0.0406 *	0.0182 *

Table 4.6: *p*-values obtained by a Wilcoxon Signed-Rank (WSR) test evaluating differences between segment duration when that segment is the only element of syllable onset or when it is part of a cluster and by GLMM modelling of the effect of the type of syllable onset (single segment or cluster) on segment durations in each of the speaker groups. *p*-values < 0.005 in the case of WSR and < 0.05 in the case of GLMM are marked by (*) and indicate significance.

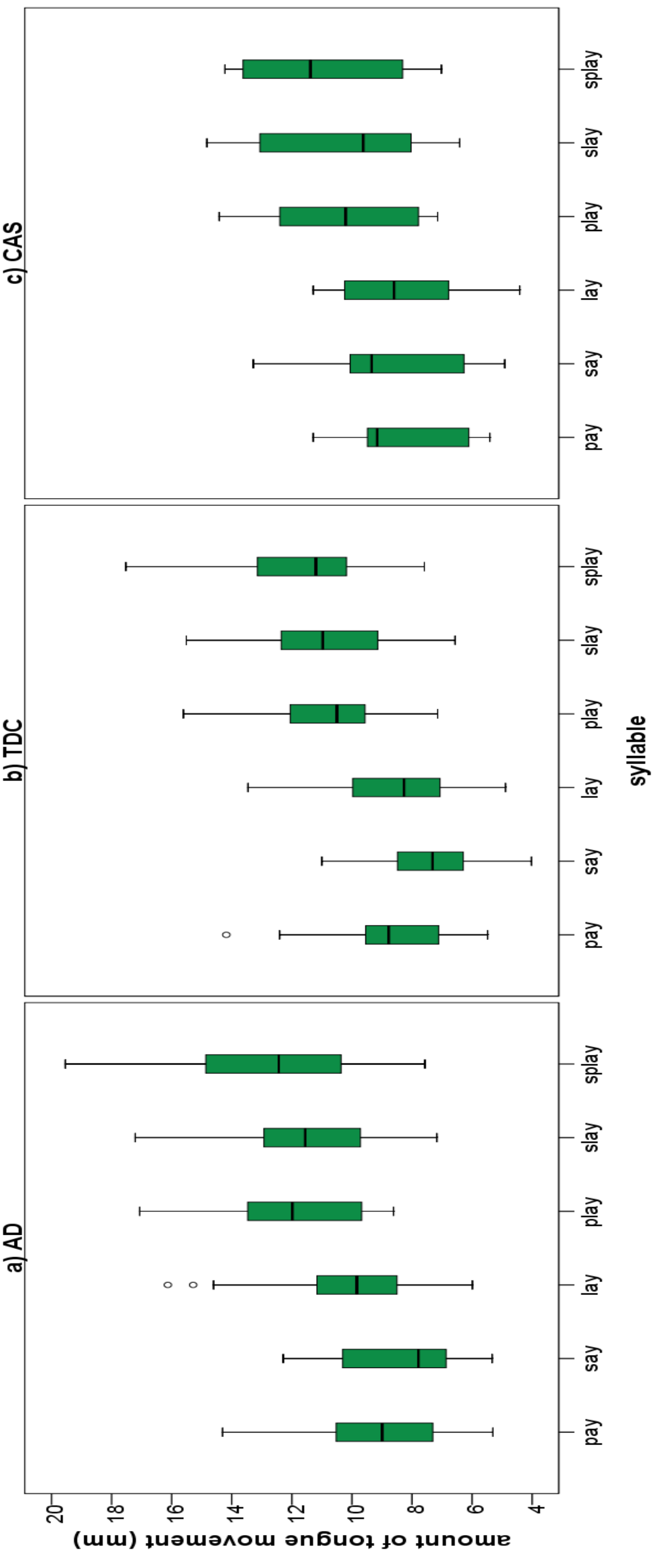


Figure 4.3: Amount of tongue movement over syllable (mm) in each of the speaker groups: (a) AD, (b) TDC, (c) CAS

Speakers	AD		TDC		CAS	
Syllable	Median	IQR	Median	IQR	Median	IQR
pay	8.99	3.26	8.77	2.46	9.16	3.43
say	7.79	3.47	7.32	2.26	9.34	3.97
lay	9.84	2.79	8.26	2.96	8.60	3.84
play	11.98	3.83	10.50	2.57	10.21	5.09
slay	11.55	3.25	10.97	3.27	9.62	5.84
splay	12.44	4.62	11.20	3.05	11.38	5.97

Table 4.7: Median and IQR values of the amount of tongue movement (mm) by syllable type and speaker group.

The significance of the increase of the amount of tongue movement with the addition of a lingual segment was first tested with a Wilcoxon Signed-Rank test and by comparing obtained p-values (Table 4.8) to the corrected alpha value of 0.008. The test revealed that the AD and TDC groups partly increased the amount of tongue movement when a lingual segment was added to the syllable onset. The AD speakers had more tongue movement over "play" than over "pay" or "lay", and more over "slay" than over "say". The TDC group showed the same result with the addition of more tongue movement over "slay" than over "lay". Speakers with CAS did not show any difference in the amount of tongue movement over tested syllable pairs. The effect of the number of lingual onset segments was additionally tested by applying GLMM. As can be seen also in Table 4.8, this method showed that number of lingual onset segments had a significant effect on the amount of tongue movement in the AD and the TDC groups but not in the CAS group.

4.2.2 Between speaker groups

Amount of tongue movement over the syllable was further inspected by comparing the measures of the same targets across speakers groups but no consistent differences between groups were noted (Figure 4.3 and Table 4.7). The CAS group had the greatest average amount of tongue movement over "pay" and "lay", and the AD group over the other four syllables. Similarly, the CAS group had the smallest amount of tongue movement over "play" and "slay", and the TDC one over the rest of the targets. The groups also differed to some extent in the achieved maximum and minimum values. Compared to the AD group, the TDC group had lower maximum and minimum values on all syllables. The CAS group showed lower maximum values than the control

	Speakers	AD	TDC	CAS
		p-values	*p*-values	*p*-values
WSR	pay-play	0.0025 *	0.0045 *	0.1090
	say-slay	0.0025 *	0.0025 *	0.1090
	lay-play	0.0035 *	0.0025 *	0.1090
	lay-slay	0.0140	0.0025 *	0.1090
	play-splay	0.0370	0.1015	0.2850
	slay-splay	0.0140	0.1425	0.1090
GLMM	Number of lingual onset segments	0.0332 *	0.0456 *	0.1800

Table 4.8: *p*-values obtained by a Wilcoxon Signed-Rank (WSR) test evaluating differences in the amount of tongue movement between syllable pairs, and by GLMM modelling of the effect of the number of lingual syllable onset segments on the amount of tongue movement over the syllable in each of the speaker groups. *p*-values < 0.008 in the case of WSR and < 0.05 in the case of GLMM are marked by (*) and indicate significance.

groups on all syllables except “say”, lower minimum values than the AD group on all syllables except “pay” and some lower (“lay”, “splay”), some higher (“say”) and some similar (“pay”, “play”, “slay”) minimum values, as compared to the TDC group. More consistent differences were observed in group variability, with the CAS group being the most variable on all of the six syllables. The AD and the TDC groups did not differ from each other so clearly as one group was more variable on some targets and the other on the rest.

However, when testing for significant differences between groups, none of the measures proved to be significant. Table 4.9 presents *p*-values obtained by a Mann-Whitney test evaluating differences in the amount of tongue movement over the syllable and its IQR across speaker groups. The alpha level was corrected to 0.008.

Similar results were revealed after applying GLMM (Table 4.9). The CAS group differed from the AD group only on one syllable, having significantly less tongue movement over “lay”. Similarly, the TDC group had significantly less tongue movement over “lay” and “say” than the AD group, but not over other syllables. The CAS and the TDC groups did not differ from each other in the amount of tongue movement over syllables. Moreover, the groups also showed the same amount of variability on almost all syllables (Table 4.10). The CAS group had lower variability than the TDC group for syllables "lay" and "splay" only, but none of these two groups differed significantly from the AD group.

Speakers	AD vs. TDC		AD vs. CAS		TDC vs. CAS	
Syllable	MW *p*-values	GLMM *p*-values	MW *p*-values	GLMM *p*-values	MW *p*-values	GLMM *p*-values
pay	0.4400	0.5354	0.8660	0.4344	1.0000	0.7196
say	0.1630	0.0000 *	0.8660	0.7636	0.3980	0.0898
lay	0.0205	0.0000 *	0.3100	0.0000 *	0.8660	0.9880
play	0.2030	0.1034	0.6120	0.0544	0.8660	0.4134
slay	0.1820	0.3286	0.3980	0.1324	0.7350	0.3842
splay	0.2030	0.1928	0.4990	0.1206	0.8660	0.4728

Table 4.9: *p*-values obtained by a Mann-Whitney (MW) test evaluating differences in the amount of tongue movement between speaker groups, and by GLMM modelling of the effect of the speaker group on the amount of tongue movement over syllable for each of the syllable types. *p*-values < 0.008 in the case of MW and < 0.05 in the case of GLMM are marked by (*) and indicate significance.

Speakers	AD vs. TDC		AD vs. CAS		TDC vs. CAS	
Syllable	MW *p*-values	GLMM *p*-values	MW *p*-values	GLMM *p*-values	MW *p*-values	GLMM *p*-values
pay	0.2480	0.5882	0.2495	0.5884	0.1990	0.4108
say	0.0755	0.1258	0.2490	0.2586	0.5000	0.9424
lay	0.2725	0.3956	0.0640	0.1430	0.0455	0.0476 *
play	0.2030	0.7674	0.5000	0.9684	0.4330	0.7872
slay	0.1130	0.3236	0.1185	0.3386	0.0215	0.1084
splay	0.3810	0.6054	0.0455	0.0974	0.0090	0.0486 *

Table 4.10: *p*-values obtained by a Mann-Whitney (MW) test evaluating differences in the IQR of the amount of tongue movement between speaker groups, and by GLMM modelling of the effect of the speaker group on the IQR of the amount of tongue movement over syllable for each of the syllable types. *p*-values < 0.008 in the case of MW and < 0.05 in the case of GLMM are marked by (*) and indicate significance.

4.3 Group results: Combining temporal and articulatory data

Another way of presenting syllable duration and tongue movement in syllable realisation is to plot these measures together, so that both distributions can be observed at the same time. Such distributions inside the speaker groups is shown in Figure 4.4. Because all plots are presented on the same scale it is possible to directly compare distribution patterns of the six syllables within and between the speaker groups and observe their characteristics. First, it can be observed that the three groups differ in the range of measured values with the TDC group showing the greatest spread. The values of the AD and the CAS groups have more similar spread, with CAS occupying a slightly narrower range, although this is likely to be due to the smaller number of speakers in the CAS group. Second, the six different syllables do not form distinctive distribution clusters in any of the speaker groups. However, although they overlap, some emergent patterns could still be noted. The AD speakers (Figure 4.4a) display similar distribution of “pay” and “say”, and of “slay” and “splay”. “lay” is more separated from the other syllables with single onset segments since it was realised with shorter duration and greater amount of tongue movement. All these syllables also show greater spread of distribution than “play”, which is positioned the most centrally in the distribution of all AD syllables. Some similar observations can be made about the TDC group distribution (Figure 4.4b). Like in the AD group, “slay” and “splay” cannot be easily distinguished from each other and “play” lies the most centrally in the distribution. However, unlike adults, “slay” and “splay” in the TDC group are more clearly separated from syllables with single onsets which are in turn more overlapped. The CAS group includes only three speakers which is likely to influence the displayed distribution (Figure 4.4c). However, it can still be observed that the distribution pattern and the relationship between syllable types do not show any clear differences from the control groups.

4.3.1 Rate of tongue movement

Acoustic and articulatory data were additionally combined by calculating the rate of tongue movement over the syllable. Figure 4.5 and Table 4.11 show group values of the rate of tongue movement over the syllable for each of the six syllables.

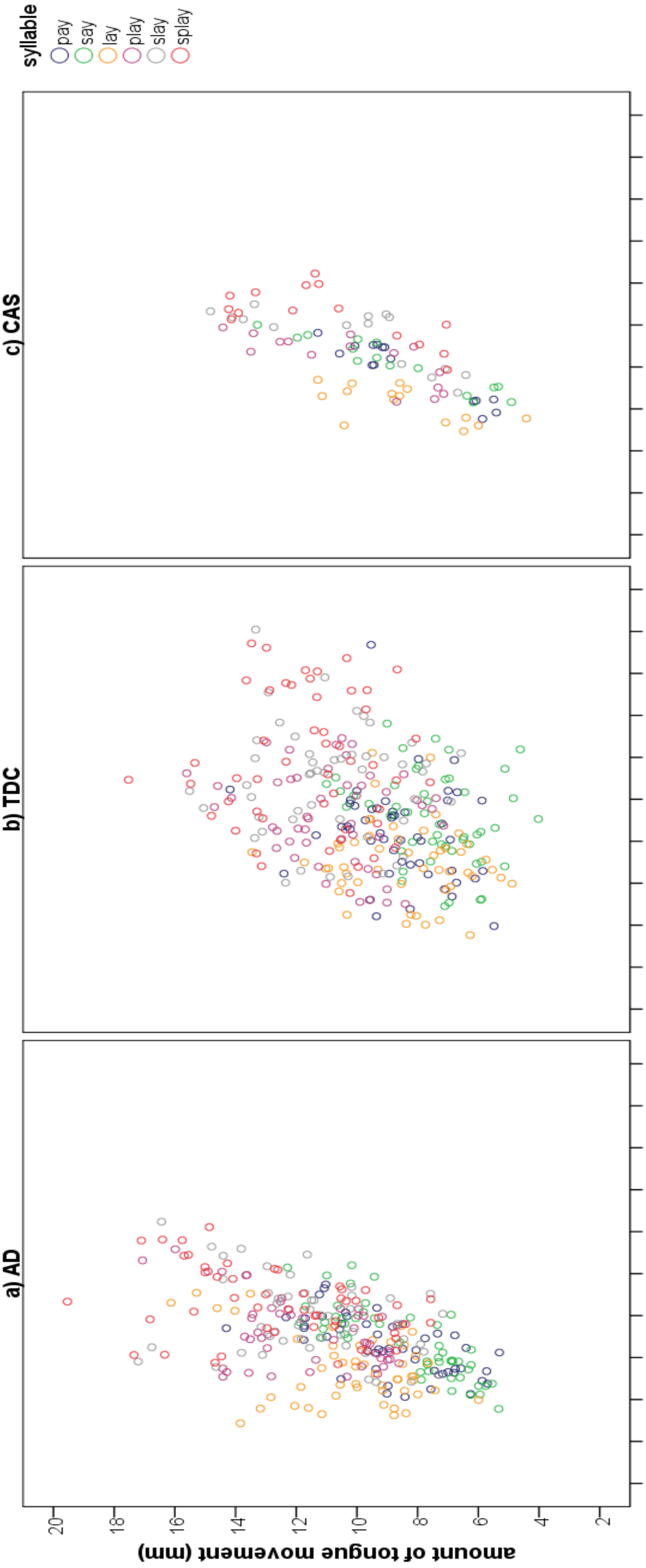


Figure 4.4: Distribution of duration (ms) and amount of tongue movement (mm) over the six syllables in each of the speaker groups: (a) AD, (b) TDC, (c) CAS.

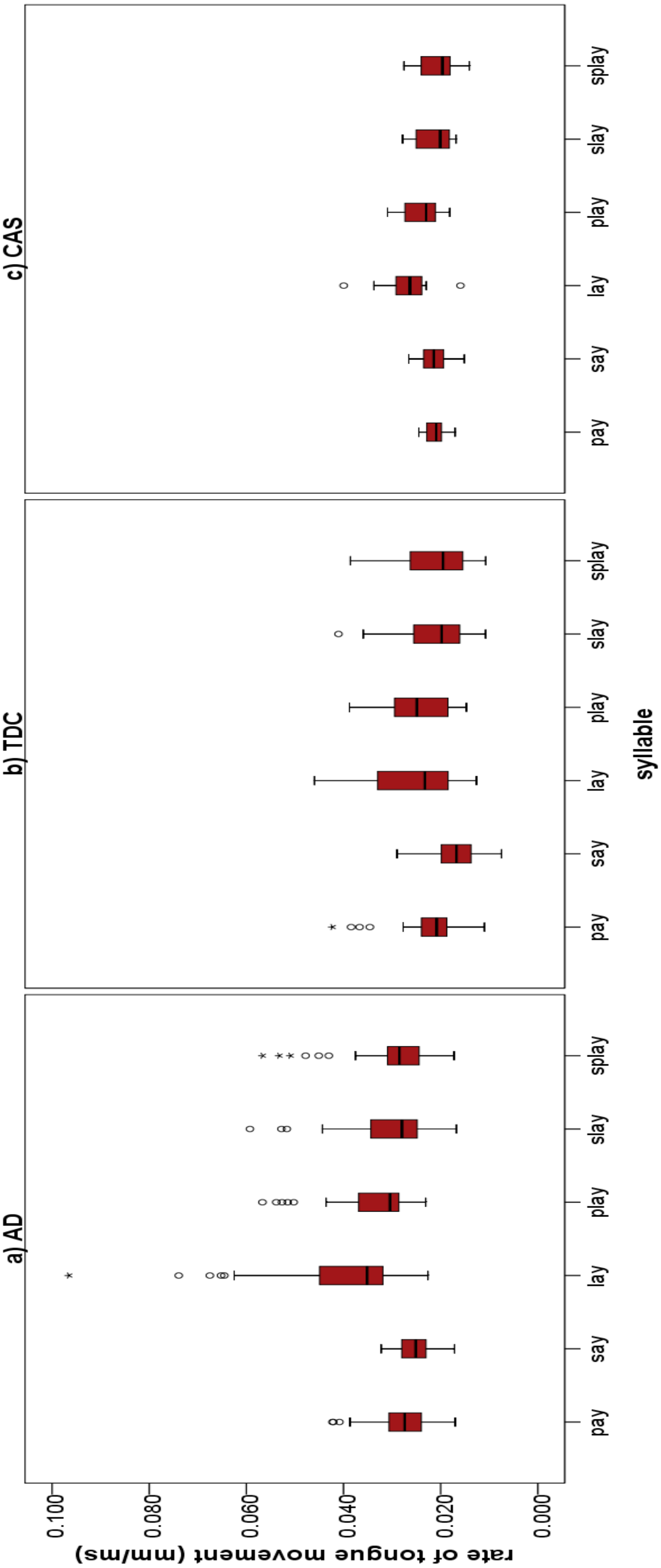


Figure 4.5: Rate of tongue movement over syllable (mm/ms) in each of the speaker groups: (a) AD, (b) TDC, (c) CAS

Speakers	AD		TDC		CAS	
Syllable	Median	IQR	Median	IQR	Median	IQR
pay	0.027	0.007	0.021	0.005	0.021	0.004
say	0.025	0.005	0.017	0.006	0.021	0.005
lay	0.035	0.014	0.023	0.015	0.026	0.006
play	0.030	0.090	0.025	0.011	0.023	0.007
slay	0.028	0.010	0.020	0.010	0.020	0.008
splay	0.029	0.007	0.020	0.011	0.020	0.007

Table 4.11: Median and IQR values of the rate of tongue movement (mm/ms) by syllable type and speaker group.

4.3.1.1 Within speaker groups

Just as with the previous measures, we were first interested in whether the number and/or type of onset segments have any significant effect on the rate of tongue movement over the syllable. As has been shown, syllable onset structure did not influence the rate of tongue movement. There is no trend of increasing or decreasing rate with the addition of onset segments, either lingual or non-lingual. The groups showed different changes in the average values for this measure over the six targets. The only common characteristic was that all three groups had the greatest rate of tongue movement over “play” and “lay”. The lack of clear effect of the number and type of onset segments was further demonstrated both by a Wilcoxon Signed-Rank test (alpha level corrected to 0.008) and GLMM (Table 4.12). A non-parametric test showed that the AD group had a higher rate of tongue movement over “lay” than over “play” and “slay”, while the TDC group had a higher rate over “lay” than over “slay” and higher over “play” than over “splay”. Speakers with CAS showed no difference in the rate of tongue movement across the tested syllable pairs. GLMM showed no effect of the number of lingual or non-lingual onset segments on the rate of tongue movement in any of the speaker groups.

4.3.1.2 Between speaker groups

Rate of tongue movement was compared between speaker groups as well. As can be seen in Figure 4.5 and Table 4.11, the AD group had the greatest rate of tongue movement over all six syllables, while similar average values were measured in the CAS and the TDC groups. This was additionally reflected when observing minimum and maximum values of the measures. All minimum and maximum values in the TDC

	Speakers	AD	TDC	CAS
		p-values	p-values	p-values
WSR	pay-play	0.0140	0.0140	0.1090
	say-slay	0.0085	0.0110	0.5930
	lay-play	0.0045 *	0.4390	0.1090
	lay-slay	0.0035 *	0.0025 *	0.1090
	play-splay	0.0110	0.0025 *	0.1090
	slay-splay	0.0845	0.2875	0.5930
GLMM	Number of lingual onset segments	0.9220	0.9174	0.4982
	Number of non-lingual onset segments	0.7644	0.5788	0.4596

Table 4.12: *p*-values obtained by a Wilcoxon Signed-Rank (WSR) test evaluating differences in the rate of tongue movement between syllable pairs and by GLMM modelling of the effect of the number of lingual and non-lingual syllable onset segments on the rate of tongue movement over the syllable in each of the speaker groups. p-values < 0.008 in the case of WSR and < 0.05 in the case of GLMM are marked by (*) and indicate significance.

group were lower than in the AD group. Lower minimum can be seen for “say”, “lay”, “play” and “splay” in the CAS group as compared to the AD one, and similar for the rest of the syllables. Like the TDC group, the CAS group also produced lower maximum values on all syllables than the AD group. Comparing the CAS and the TDC groups showed that measures in the CAS group overlap with those in the TDC, as none of the extreme values were outside of the TDC range. The presented data also shows smaller within-group variability in the CAS group and no clear differences between the control groups. These observations were further explored by applying statistical test to reveal their significance.

Firstly, a Mann-Whitney test (Tables 4.13 and 4.14) with alpha level corrected to 0.003 revealed only very small differences between the groups. The AD group had greater rate of tongue movement over syllables with single segment onsets than the TDC group. However, the TDC group showed significantly less within-group variability in “pay” than the AD group. No other differences in variability proved to be significant. CAS speakers did not differ from the control groups in the rate of tongue movement or in the variability of this measure.

However, following GLMM (Table 4.13), a significant (p-value < 0.05) effect was shown of the speaker group on all syllable targets when comparing the AD to the TDC group, and the AD to the CAS group. Both the TDC and the CAS groups had significantly slower rates of tongue movement than the AD. The CAS and the TDC groups differed only on “say” where the TDC one had slower rates of movements. The three groups showed almost the same amount of variability with the exception of the TDC and the CAS groups being less variable over “pay” than the AD group (Table 4.14). No other differences were noted on the rest of the syllables.

4.4 Patterns of tongue movement in individual speakers

In addition to measuring different aspects of tongue movement over the syllable, an attempt was made to describe them qualitatively as well. This was achieved by observing the changes in the whole midsagittal tongue contour over the syllable and changes in the highest point of the midsagittal tongue contour.

Because of the amount of collected data and differences in the quality of the obtained ultrasound image across speakers, qualitative analysis was applied only to the

Speakers	AD vs. TDC		AD vs. CAS		TDC vs. CAS	
Syllable	MW *p*-values	GLMM *p*-values	MW *p*-values	GLMM *p*-values	MW *p*-values	GLMM *p*-values
pay	0.0010 *	0.0016 *	0.0110	0.0048 *	1.0000	0.8896
say	0.0010 *	0.0002 *	0.1280	0.0168 *	0.2370	0.0116 *
lay	0.0025 *	0.0002 *	0.0430	0.0010 *	0.7350	0.5650
play	0.0115	0.0010 *	0.0180	0.0010 *	0.6120	0.9224
slay	0.0080	0.0016 *	0.0630	0.0034 *	0.6120	0.9260
splay	0.0065	0.0001 *	0.0430	0.0014 *	0.6120	0.9792

Table 4.13: *p*-values obtained by a Mann-Whitney (MW) test evaluating differences in the rate of tongue movement between speaker groups and by GLMM modelling of the effect of the speaker group on the rate of tongue movements over the syllable for each of the syllable types. *p*-values < 0.008 in the case of MW and < 0.05 in the case of GLMM are marked by (*) and indicate significance.

Speakers	AD vs. TDC		AD vs. CAS		TDC vs. CAS	
Syllable	MW *p*-values	GLMM *p*-values	MW *p*-values	GLMM *p*-values	MW *p*-values	GLMM *p*-values
pay	0.0041 *	0.0732	0.0315	0.0000 *	0.1550	0.3578
say	0.4105	0.9284	0.3675	0.7182	0.3675	0.7804
lay	0.4400	0.7172	0.2495	0.6052	0.1185	0.4552
play	0.4400	0.6380	0.4330	0.6272	0.4330	0.8720
slay	0.3810	0.3614	0.4330	0.4442	0.4330	0.8918
splay	0.4105	0.7376	0.0880	0.1372	0.0455	0.1988

Table 4.14: *p*-values obtained by a Mann-Whitney (MW) test evaluating differences in the IQR of the rate of tongue movement between speaker groups and by GLMM modelling of the effect of the speaker group on the IQR of the rate of tongue movements over the syllable for each of the syllable types. *p*-values < 0.008 in the case of MW and < 0.05 in the case of GLMM are marked by (*) and indicate significance.

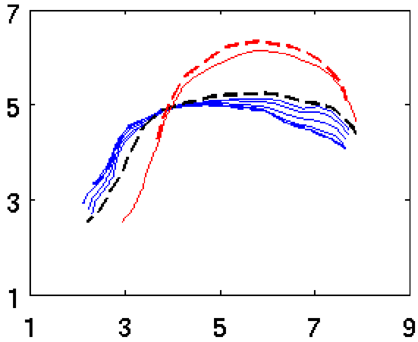
three speakers with CAS, and three speakers in each of the control groups. The AD and the TDC speakers were selected on the basis of the amount of tongue contour visible on the ultrasound images (potential causes of poor image quality are discussed in Section 2.4). The three speakers with the greatest portion of the tongue visible on all of the scanned frames were chosen as described in Section 3.4.4. The selected AD speakers were AD3, AD5 and AD9, and the selected TDC speakers were TDC1, TDC9 and TDC10.

4.4.1 Changes in midsagittal tongue contour over the syllable

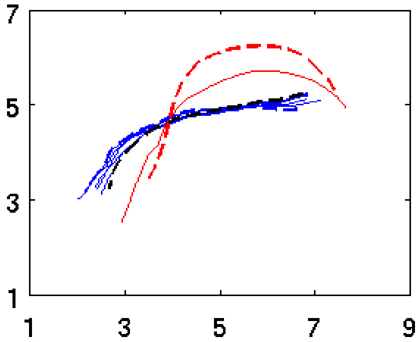
Since ultrasound enables imaging the whole midsagittal tongue contour it is appropriate for observing changes in tongue movement over different syllables and across speakers even though it cannot image the raised tongue tip. This limitation can prevent assessment of the exact position of the tongue tip and the length of the tongue, which makes it unreliable to divide the tongue into exact anatomic sections such as tongue tip, front, middle and back. However, the greatest portion of the tongue is still visible and that allows describing tongue patterns in terms of the horizontal or vertical movements of the entire tongue or of more loosely defined front, middle and back regions of the scanned image.

Figures 6-14 present midsagittal tongue contours of one repetition of the six syllables for each of the selected speakers. The AD speakers are presented in Figure 4.6 (speaker AD1), Figure 4.7 (speaker AD5), and Figure 4.8 (speaker AD9), the TDC speakers in Figure 4.9 (speaker TDC1), Figure 4.10 (speaker TDC9), and Figure 4.11 (speakers TDC10), and, finally, the CAS speakers in Figure 4.12 (speaker CAS1), Figure 4.13 (speaker CAS2), and Figure 4.14 (speaker CAS3). Each set of tongue contours representing tongue movement over the syllable is colour-coded in the following way: a dashed blue contour represents the first tongue position of the syllable, solid blue contours represent the tongue contours over syllable onset, a dashed black contour is the contour at the boundary between onset and following vowel, a solid red contour shows the tongue contour in the middle of the vowel and a dashed red contour the tongue position at the end of the syllable. Midsagittal contours of all repetitions by nine speakers can be seen below in Figures 4.15 - 4.41.

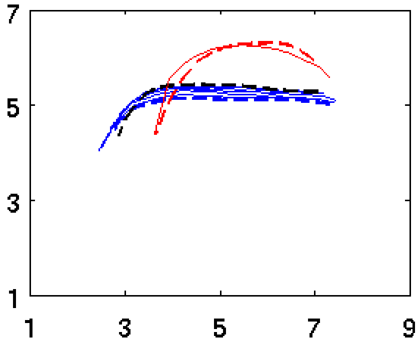
As can be seen in Figures 4.6, 4.7 and 4.8, the three AD speakers showed similar tongue movement patterns over the syllable. First, it is interesting to observe the tongue contours during the articulation of the vowel /e/. All speakers produced the expected



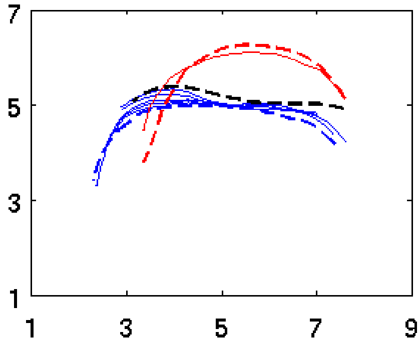
(a) pay



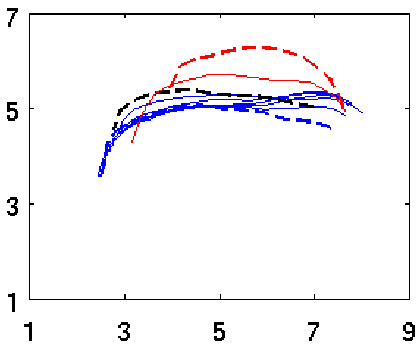
(b) say



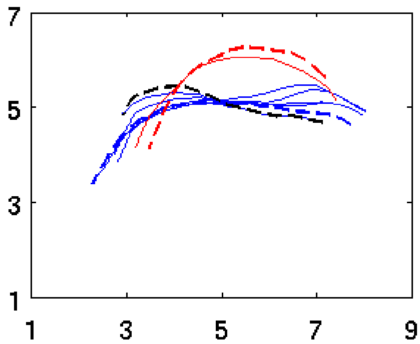
(c) lay



(d) play

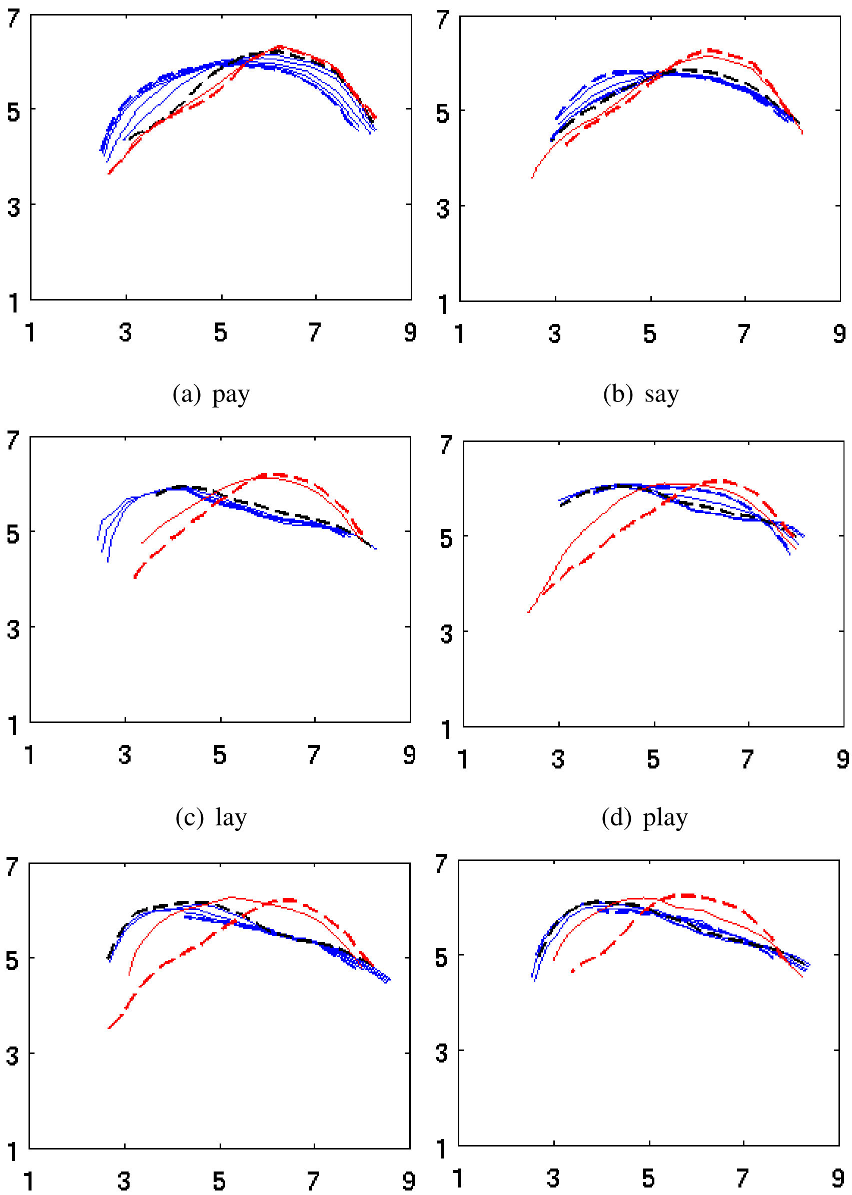


(e) slay



(f) splay

Figure 4.6: Speaker AD1's midsagittal tongue contours over the syllables: (a) pay, (b) say, (c) lay, (d) play, (e) slay, (f) splay. Contours are colour coded in the following way: dashed blue contour = the first tongue position of the syllable, solid blue contours = five tongue contours over syllable onset, dashed black contour = the boundary between onset and following vowel, solid red contour = the middle of the vowel, dashed red contour = the end of the syllable. Scale is in cm.

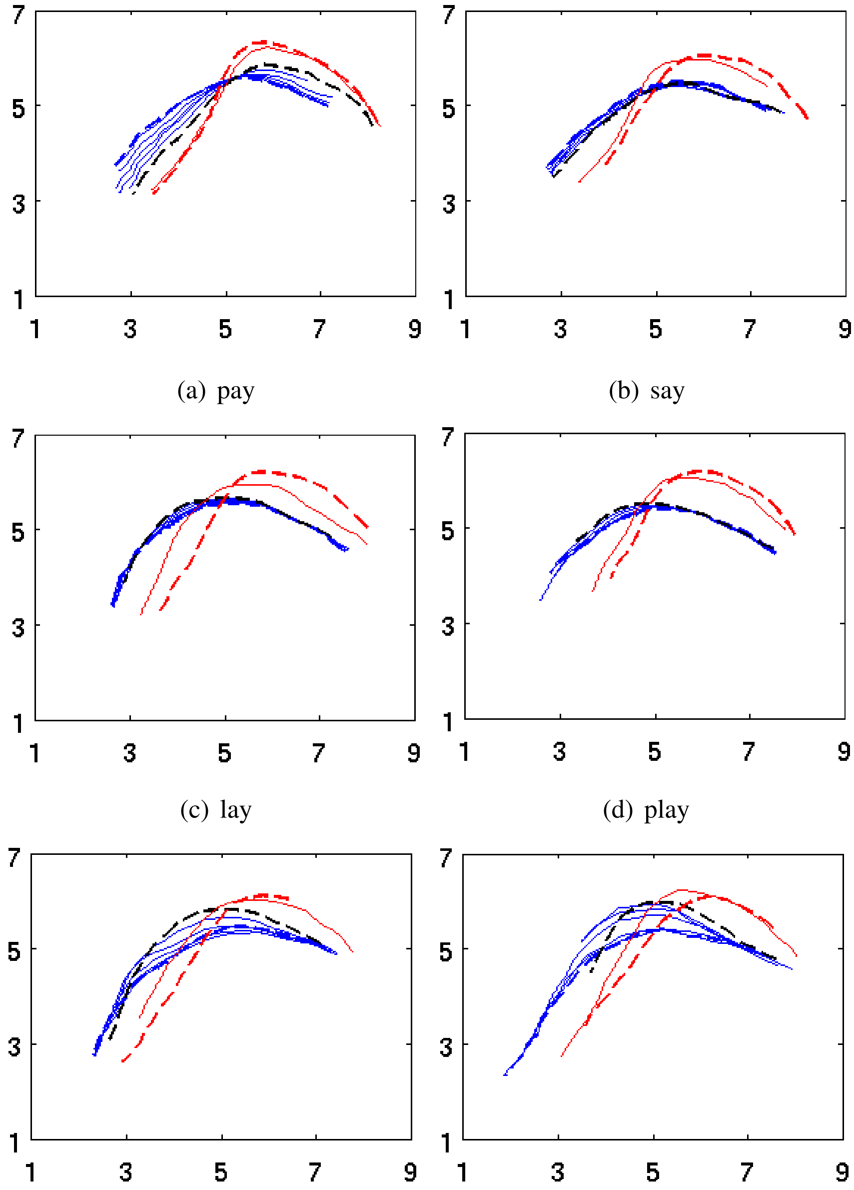


(a) pay (b) say

(c) lay (d) play

(e) slay (f) splay

Figure 4.7: Speaker AD5's midsagittal tongue contours over the syllables: (a) pay, (b) say, (c) lay, (d) play, (e) slay, (f) splay. Contours are colour coded in the following way: dashed blue contour = the first tongue position of the syllable, solid blue contours = five tongue contours over syllable onset, dashed black contour = the boundary between onset and following vowel, solid red contour = the middle of the vowel, dashed red contour = the end of the syllable. Scale is in cm.

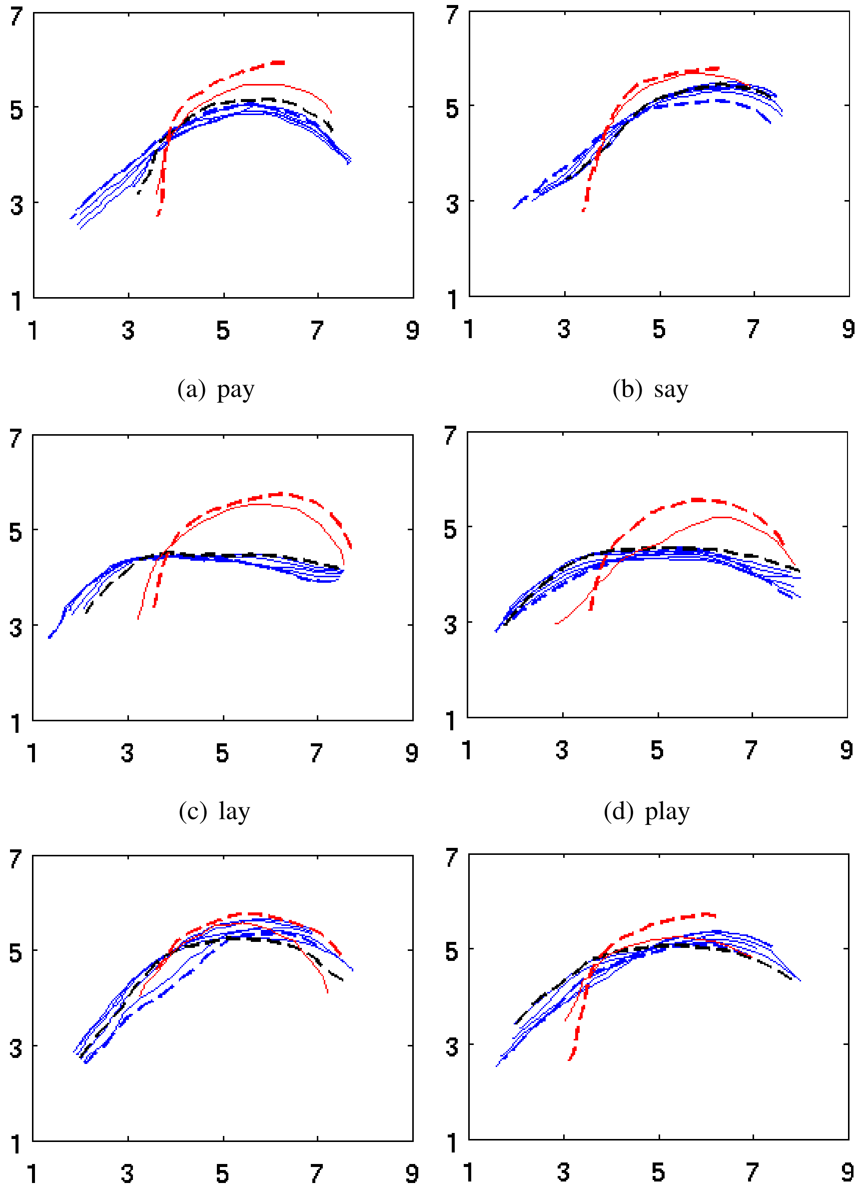


(a) pay (b) say

(c) lay (d) play

(e) slay (f) splay

Figure 4.8: Speaker AD9's midsagittal tongue contours over the syllables: (a) pay, (b) say, (c) lay, (d) play, (e) slay, (f) splay. Contours are colour coded in the following way: dashed blue contour = the first tongue position of the syllable, solid blue contours = five tongue contours over syllable onset, dashed black contour = the boundary between onset and following vowel, solid red contour = the middle of the vowel, dashed red contour = the end of the syllable. Scale is in cm.



(a) pay (b) say

(c) lay (d) play

(e) slay (f) splay

Figure 4.9: Speaker TDC1's midsagittal tongue contours over the syllables: (a) pay, (b) say, (c) lay, (d) play, (e) slay, (f) splay. Contours are colour coded in the following way: dashed blue contour = the first tongue position of the syllable, solid blue contours = five tongue contours over syllable onset, dashed black contour = the boundary between onset and following vowel, solid red contour = the middle of the vowel, dashed red contour = the end of the syllable. Scale is in cm.

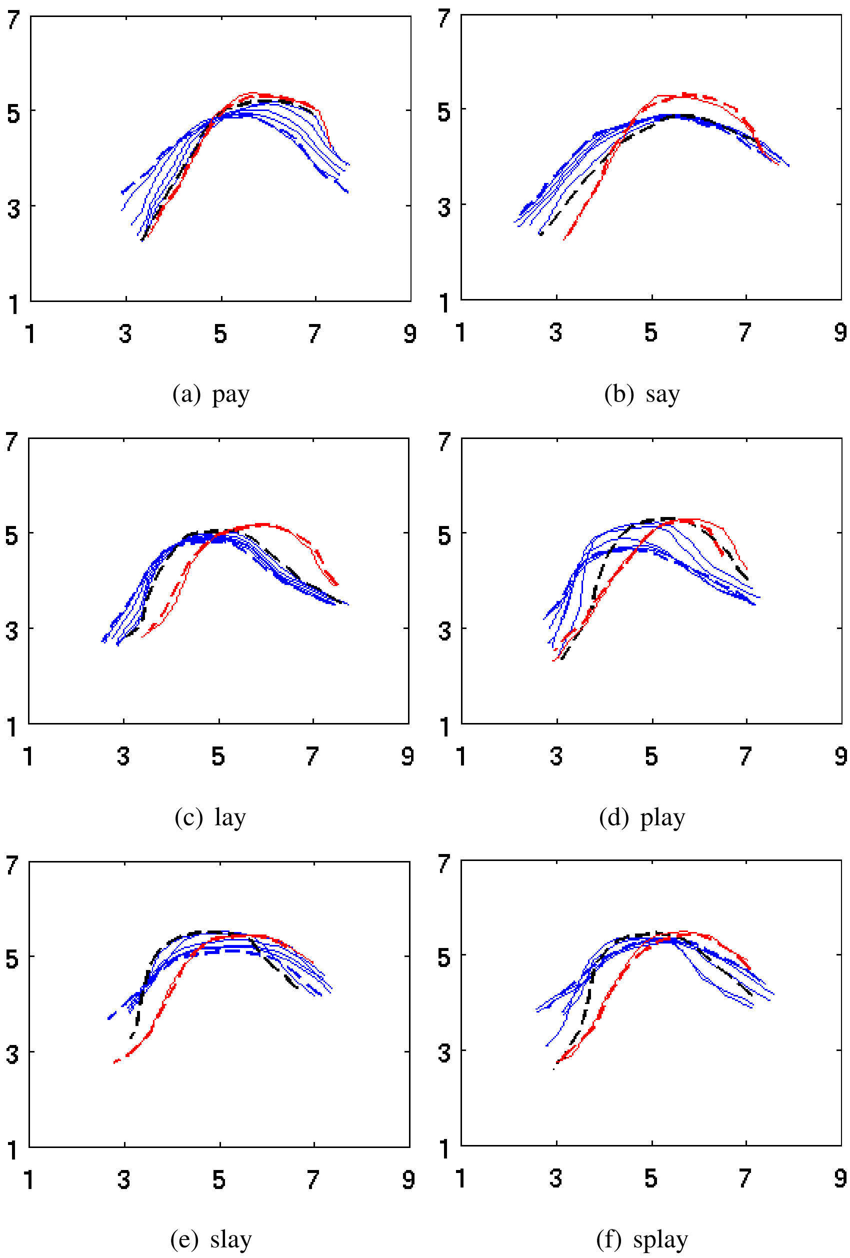
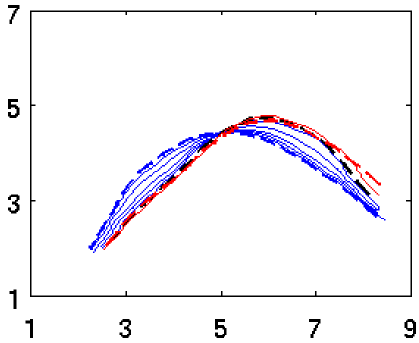
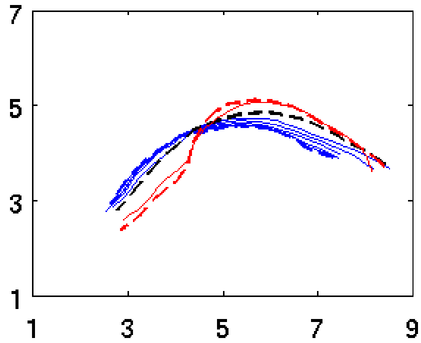


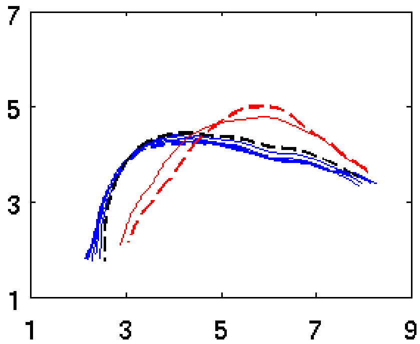
Figure 4.10: Speaker TDC9's midsagittal tongue contours over the syllables: (a) pay, (b) say, (c) lay, (d) play, (e) slay, (f) splay. Contours are colour coded in the following way: dashed blue contour = the first tongue position of the syllable, solid blue contours = five tongue contours over syllable onset, dashed black contour = the boundary between onset and following vowel, solid red contour = the middle of the vowel, dashed red contour = the end of the syllable. Scale is in cm.



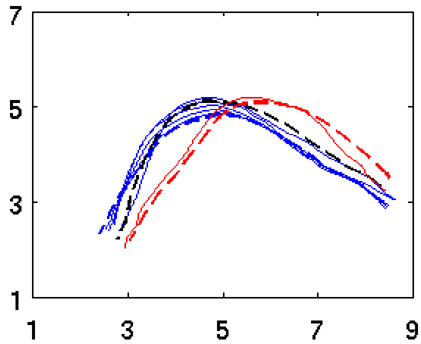
(a) pay



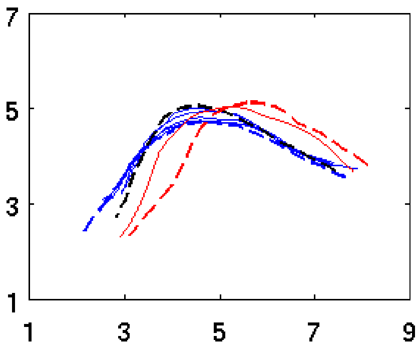
(b) say



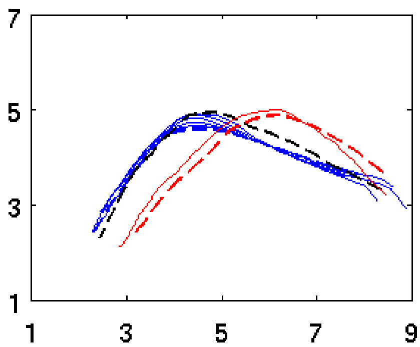
(c) lay



(d) play

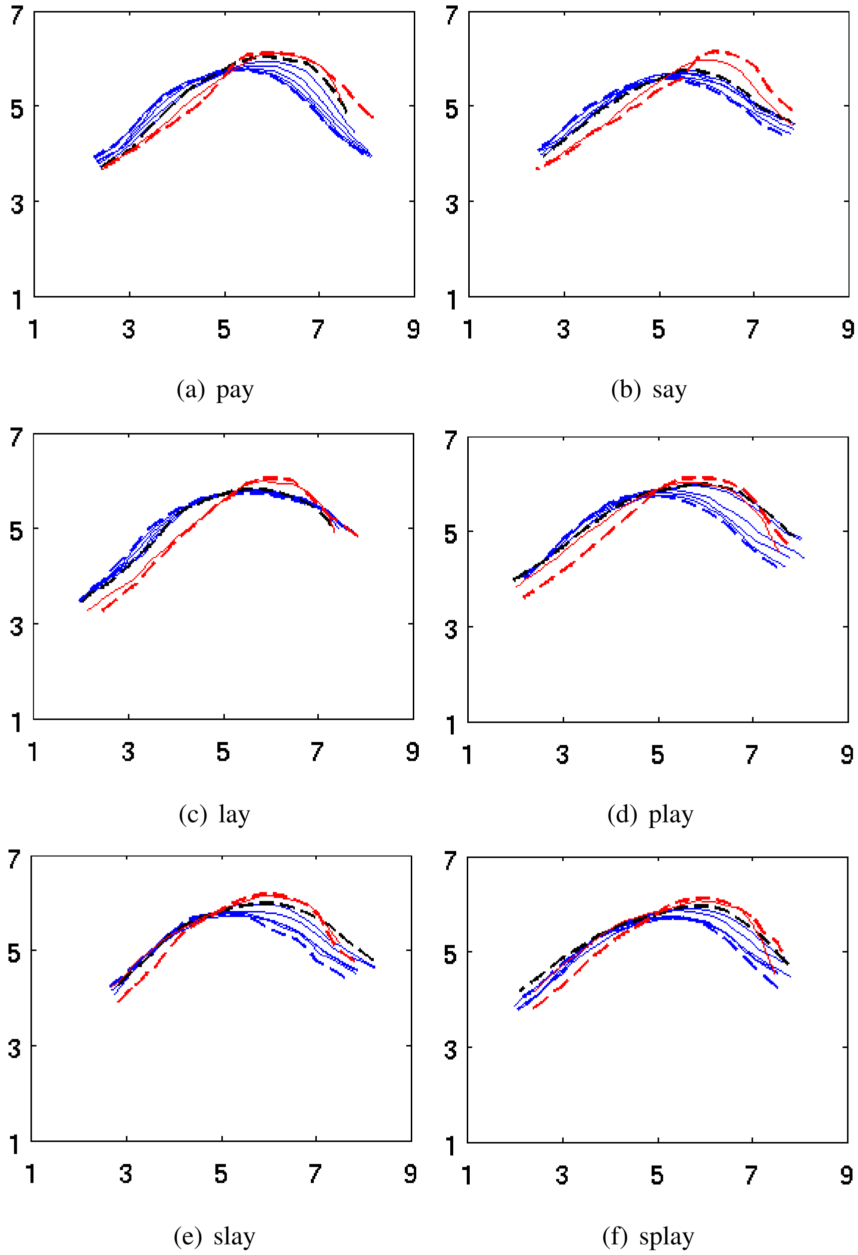


(e) slay



(f) splay

Figure 4.11: Speaker TDC10's midsagittal tongue contours over the syllables: (a) pay, (b) say, (c) lay, (d) play, (e) slay, (f) splay. Contours are colour coded in the following way: dashed blue contour = the first tongue position of the syllable, solid blue contours = five tongue contours over syllable onset, dashed black contour = the boundary between onset and following vowel, solid red contour = the middle of the vowel, dashed red contour = the end of the syllable. Scale is in cm.

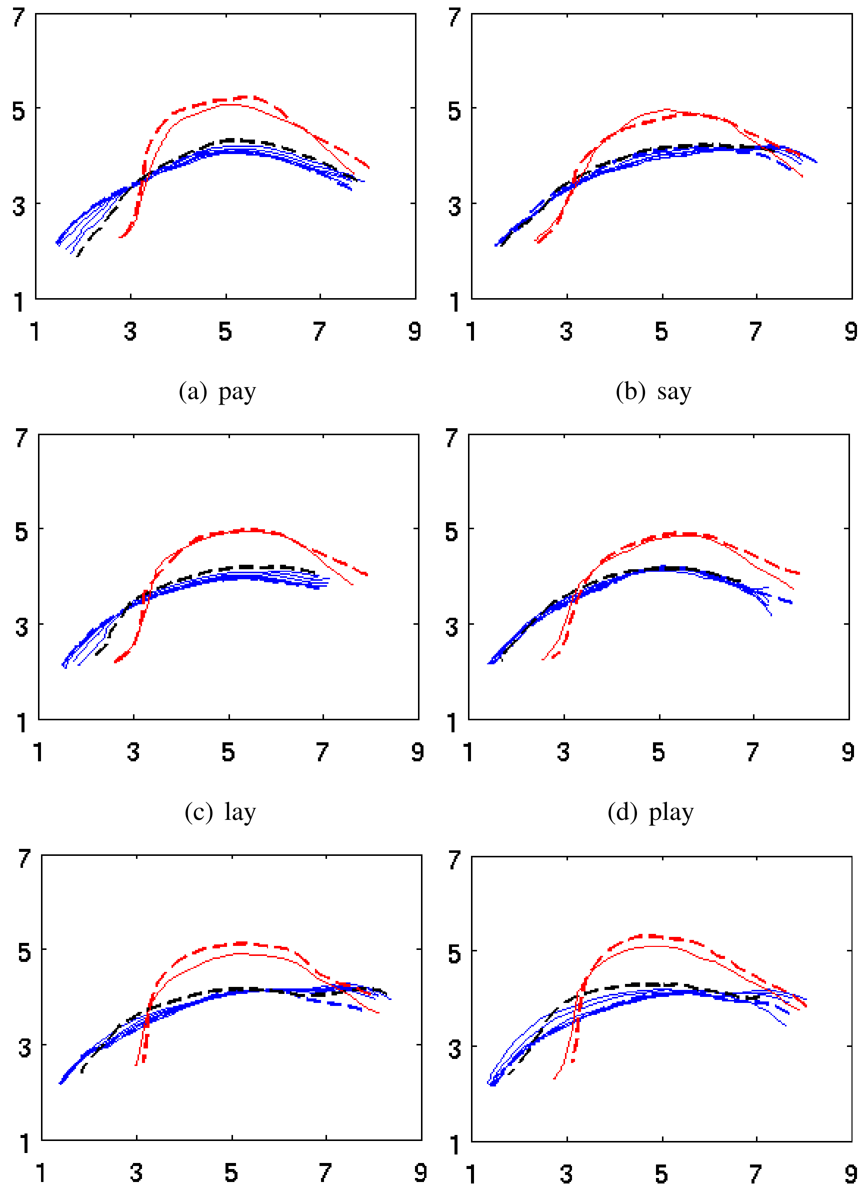


(a) pay (b) say

(c) lay (d) play

(e) slay (f) splay

Figure 4.12: Speaker CAS1's midsagittal tongue contours over the syllables: (a) pay, (b) say, (c) lay, (d) play, (e) slay, (f) splay. Contours are colour coded in the following way: dashed blue contour = the first tongue position of the syllable, solid blue contours = five tongue contours over syllable onset, dashed black contour = the boundary between onset and following vowel, solid red contour = the middle of the vowel, dashed red contour = the end of the syllable. Scale is in cm.

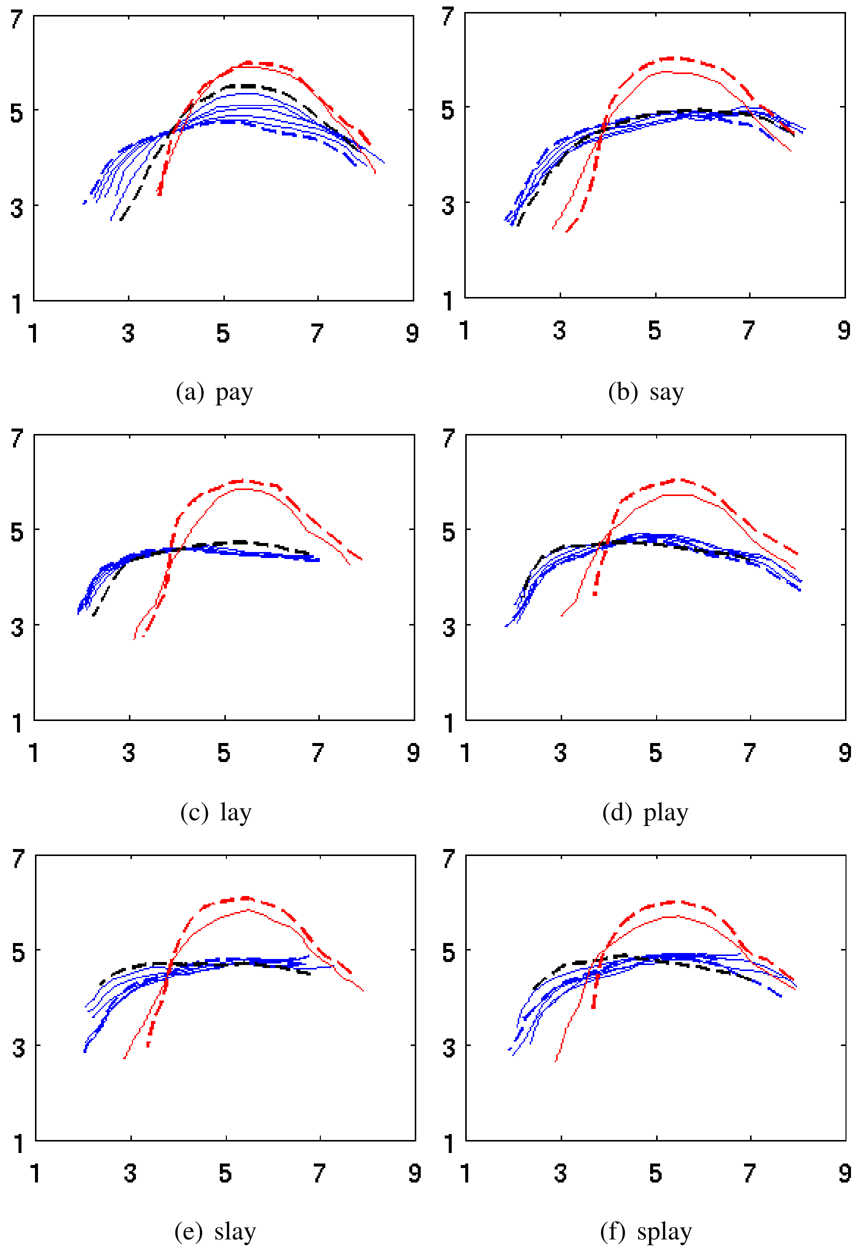


(a) pay (b) say

(c) lay (d) play

(e) slay (f) splay

Figure 4.13: Speaker CAS2's midsagittal tongue contours over the syllables: (a) pay, (b) say, (c) lay, (d) play, (e) slay, (f) splay. Contours are colour coded in the following way: dashed blue contour = the first tongue position of the syllable, solid blue contours = five tongue contours over syllable onset, dashed black contour = the boundary between onset and following vowel, solid red contour = the middle of the vowel, dashed red contour = the end of the syllable. Scale is in cm.



(a) pay (b) say (c) lay (d) play (e) slay (f) splay

Figure 4.14: Speaker CAS3's midsagittal tongue contours over the syllables: (a) pay, (b) say, (c) lay, (d) play, (e) slay, (f) splay. Contours are colour coded in the following way: dashed blue contour = the first tongue position of the syllable, solid blue contours = five tongue contours over syllable onset, dashed black contour = the boundary between onset and following vowel, solid red contour = the middle of the vowel, dashed red contour = the end of the syllable. Scale is in cm.

shape of raised front of the tongue and individual speakers showed similar tongue shapes in different syllables. This was not surprising as only the preceding consonant was changing and the vowel was always followed by /t/. Comparing tongue position in /e/ across three AD speakers also revealed some differences in the height of the tongue. For example, speaker AD5 (Figure 4.7) had less upward movement from the onset to the vowel in all syllables than speakers AD3 and AD9.

Similarities were observed in the tongue movements over the syllable onsets as well. All three speakers articulated "pay" by first moving the back of the tongue forward and the front upwards over the onset /p/ and then raising the front part of the tongue high up to produce /e/ (Figures 4.6a, 4.7a and 4.8a). Because /p/ is not a lingual consonant, the movement over the onset reflected the necessary movement from preceding /ə/ to the following /e/. The movement from /ə/ to the first consonant of the syllable onset is not visible in the syllables starting with lingual /s/ and /l/. During /s/ in "say" the tongue stayed almost static with only slight movement downwards at the back of the tongue and upwards at the front of the tongue (Figures 4.6b, 4.7b and 4.8b). All three AD speakers articulated /l/ in "lay" (Figures 4.6c, 4.7c and 4.8c) by raising the back of the tongue and lowering the front of the tongue already in the first frame of the onset and keeping the tongue relatively static during /l/, before moving the front of the tongue forward and up for /e/. Some differences between speakers can be seen in the height of the back of the tongue during /l/. Speakers AD5 and AD9 had higher tongue position, and for that reason a more curved shape of the entire tongue than AD3.

More complex tongue movement patterns were expected to be observed over clustered syllable onsets. However, inspecting the tongue movement patterns in "play", "slay" and "splay" (Figures 4.6d-f, 4.7d-f and 4.8d-f) revealed that they look very similar to tongue movements over "lay". /l/ was the only consonant with a raised back of the tongue and this upward movement started early during the articulation of the onset cluster. In addition to back raising, /pl/ is characterised by moving from a curved tongue contour positioned in the centre of the oral cavity for realising /ə/ into a raised back and lower front for /l/ in speakers AD3 and AD5. Speaker AD9, in contrast, showed an almost static tongue over the cluster with very little movement in the back tongue region. AD9 also showed more curved tongue contours in all three clustered onsets than the other two speakers who showed raised back of the tongue, lower front and a flatter middle part. A possible explanation for this is that AD9 had less of the tongue tip visible in the scanned images than AD3 and AD9, and for that reason the movement of the front of the tongue cannot be observed and the tongue shape looks more curved. Some differences were present in the articulation of /sl/ as well. Speakers AD3 and AD9 raised the back of the tongue during /l/, while speaker AD5 kept the back of the tongue in the same position during the cluster. AD3 additionally lifted the front of the tongue during /s/ and lowered it during /l/. The same was not observed for the other two speakers, most likely due to the missing raised tongue tip on the ultrasound images. Almost identical tongue movements as over /sl/ were observed also over /spl/ for all three speakers. This is not surprising since /p/ does not contribute to the movement but its presence only adds time during which /s/ and /l/ can be articulated.

Tongue movement patterns of TDC speakers are presented in Figures 4.9, 4.10 and 4.11 and, interestingly, do not differ from the tongue patterns of AD speakers. Contours representing the vowel again show the expected curve with a raised front part of the tongue, and like the AD speakers, individual TDC participants showed similar contours in different syllables. Different height and slightly different shape of the vowel contours across speakers probably reflect anatomical differences in the shape of the palate.

TDC showed the same patterns as the AD speakers over syllable onsets. /p/ in "pay" is characterised by moving the back of the tongue forward and the front up (Figures 4.9a, 4.10a and 4.11a). During /s/ in "say" (Figures 4.9b, 4.10b and 4.11b) all three speakers lowered the back of the tongue and TDC1 and TDC10 rose at the front. The same rise is not observed for TDC9, most likely due to missing tongue tip on the scanned images. More differences between the speakers are observed in the realisation of /l/ in "lay" (Figures 4.9c, 4.10c and 4.11c). TDC9 and TDC10 raised the back of the tongue but TDC9 has a lower position of the front of the tongue, probably because of a smaller oral cavity and higher palate. TDC1, on the other hand, had almost no tongue movement in the articulation of /l/. The tongue is curved, with the centre raised, throughout the consonant.

As with the AD speakers, tongue movement patterns over clusters looked very similar to those over /l/ for TDC as well. TDC9 and TDC10 both had the back of the tongue positioned higher than TDC1 in all syllables with clustered onsets. However, all three TDC raised both back and front of the tongue over /pl/ in "play" to achieve the target tongue position for the articulation of /l/ (Figures 4.9d, 4.10d and 4.11d). They also moved the back of the tongue up during the articulation of /l/ in "slay" (Figures 4.9e, 4.10e and 4.11e). In the same onset TDC1 and TDC9 lifted the front of the tongue during /s/ and lowered it again for /l/. The same was not observed for speaker

TDC10 who had no movement in the front part of the visible tongue image during /sl/. Similarly to AD, TDC showed almost identical tongue movement patterns over /spl/ as over /sl/ (Figures 4.9f, 4.10f and 4.11f), with only slightly more movement upwards in the front part of the tongue for TDC10.

Finally, tongue movement patterns for speakers with CAS can be seen in Figures 4.12, 4.13 and 4.14. Comparing the tongue contours of the vowel shows the first striking difference between CAS and both AD and TDC. /e/ is typically articulated in the front of the oral cavity and with a raised front of the tongue but CAS speakers do not show this same tongue position. As can be seen in all syllables of CAS2 (Figure 4.13) and CAS3 (Figure 4.14), and in “splay” of CAS1 (Figure 4.12f), the vowel is articulated with a high centre or even back (CAS2 in “lay” and “splay”) part of the tongue.

Looking at tongue movement over syllable onset revealed that CAS speakers showed similar patterns over /p/ in “pay” and /s/ in “say” to those of AD and TDC. In the first case (Figures 4.12a, 4.13a and 4.14a) the back of the tongue moved forward and the front up, and in the second (Figures 4.12b, 4.13b and 4.14b), the back stayed almost static and little raising was observed in the front part of the tongue. Different movement than in the control groups was again observed for the articulation of /l/ in “lay” and all the clustered onsets. In contrast to AD and TDC speakers, CAS1 and CAS2 (with the exception of “splay”) lacked raising of the back of the tongue which is typical for the articulation of /l/. The back of the tongue was instead curved downwards and lower than the front. Additionally, they also raised the front of the tongue during /l/. CAS3, on the other hand, did show the raising of the back and lowering of the front of the tongue in all of these onsets. CAS1 showed almost identical patterns of tongue movement over clustered onsets (Figure 4.12d-f). The tongue was curved with lowered and almost static back, and lowered front which moved upwards during /s/ and /l/. More movement in the front of the tongue could be observed for speaker CAS2, particularly in “slay” and “splay”. For this speaker the front of the tongue was higher than the back, rising into /s/, and in the case of /sl/ staying static during both segments (Figure 4.13e), while lowering during /l/ in /spl/ (Figure 4.13f).

4.4.2 Analysis of tongue contour patterns occurring in transitions between segments

Another attempt to describe tongue movement patterns was by applying methodology by Iskarous (2005) (details are given in Section 2.3.1), who described tongue movements by describing tongue contour patterns occurring in transitions between segments. All transitions between segments were inspected and assigned either a pivot, arch or combined pivot/arch pattern. The resulting patterns are presented in Table 4.15, for one repetition of the six syllables of each of the selected nine speakers.

Syllable	pay	say	lay	play	slay	splay
Speaker	/p/-/e/	/s/-/e/	/l/-/e/	/p/-/l/-/e/	/s/-/l/-/e/	/s/-/p/-/l/-/e/
AD3	P	P	P	A-P	P-P	A-P-P
AD5	P	P	P	A/P-P	A-P	A-A-P
AD9	P	P	P	A-P	A/P-P	A-A-P
TDC1	P	P	P	A-P	A-P	A-P-P
TDC9	P	P	P	P-P	A/P-P	P-A-P
TDC10	P	P	P	A-P	A/P-P	A-A-P
CAS1	P	P	P	A-P	A-P	A-A-P
CAS2	P	P	P	A-P	A-P	P-P-P
CAS3	P	P	P	A-P	A-P	P-P-P

Table 4.15: Type of midsagittal tongue contour transitions between syllable segments for three AD, TDC and CAS speakers. A = arch pattern transition, P = pivot pattern transition, P/A = pivot/arch pattern transition.

As can be seen, all transitions between a single segment onset and the following vowel make a pivot pattern. The pivot pattern was observed also in all transitions between /l/ and /e/ in syllables with onset clusters. The transition pattern between onset and vowel is visible in the difference between the blue contours (representing onset), the black contour (traced at the acoustic onset-vowel boundary) and red contours (representing the vowel) in Figures 4.6 - 4.14. More differences were observed in transitions between the clustered consonants (patterns emerging from the blue contours in these figures), but there are no clear distinctions between AD, TDC and CAS speakers. The transition between /p/ and /l/ in "play" forms an arch pattern in most speakers, but not in AD5 where it is a combination of arch and pivot and in TDC9 where it forms a pivot. An arch pattern was observed also in all CAS /s/ - /l/ transitions in "slay", but only in one AD and one TDC speaker. Two of the TDC speakers and one AD had a combined arch/pivot pattern, and the remaining AD a pivot transition. Different transition patterns were found between consonants in "splay" as well. Three AD speakers, two TDC speakers and one CAS speaker showed an arch pattern in /s/ - /p/ transition, and the rest a pivot pattern. The /p/ - /l/ transition resulted in an arch pattern for two AD, two TDC, and one CAS speaker, and pivot pattern for the rest. CAS2 and CAS3 had only the pivot pattern in all "splay" transitions.

4.4.3 Consistency of tongue movement across repetitions

Plotting together all tongue contours of each syllable additionally enabled observation of how consistent each individual speaker's tongue movement was over repetitions. Because of the high number of figures, the tongue contour patterns of all repetitions of the nine selected speakers are presented in Figures 4.15 - 4.41. As can be observed, all speakers have very consistent patterns over repetitions. Some more considerable inconsistencies could be observed only in one repetition of "say" (Figure 4.39) and two of "play" (Figure 4.40) uttered by CAS3, and three repetitions of "lay" by TDC1 (Figure 4.25). They all show more upward tongue movement over the onset, especially in the back part of the tongue.

Another way of exploring the consistency of tongue movements was achieved by focusing only on one aspect of tongue movement: the highest point on the tongue. After the highest point was identified on every contour in a syllable, they were all plotted together to present the highest point movement pattern. Resulting plots for one AD speaker (AD5), one TDC speaker (TDC1) and the three speakers with CAS can be seen in Figures 4.42 - 4.56 and plots of remaining two AD and two TDC in Figures 1 - 12 in Appendix III. The highest point on the first contour is always represented by a blue asterisk and the highest point on the last contour by the red one. The black line connects the highest points (represented by circles) in temporal order, from first to last.

It is important to stress that the figures do not represent any kind of point tracking, but that every circle in the figures represents the highest point of the tongue in each of the scanned frames. Because of the many possible directions of tongue movement, it is not possible to speculate on the basis of these figures about which part of the tongue is the highest at a particular time in the syllable. In fact, these figures tell us more about areas of the oral cavity that are occupied by the highest part of the tongue than about which part of the tongue is the highest. Although it is possible that the same part of the tongue has the highest position in the same area of the oral cavity in more than one frame, it is more likely that the highest part of the tongue and/or the occupied area change due to whole tongue movement and a change in shape. For example, at the beginning of articulation of “pay”, the middle part of the tongue has the highest point but towards the end it is the front part, which has to be lifted to achieve the correct shape for /e/.

Presentation of the highest point on the tongue shows more differences between the speakers. The differences are, however, not the expected consistent repetitions in the AD group and more varied repetitions of the TDC and the CAS speakers. Figures 4.42 - 4.44 and 4.48 - 4.56 clearly show that the AD and the CAS speakers have consistent patterns of the highest point movement in the repetitions. Patterns of the remaining two selected AD speakers presented in Appendix III show the same consistency of repetition with the exception of ”lay” by AD3 (Appendix III, Figure 2) where two different patterns of movement can be seen. In contrast, TDC speakers articulated more syllables with different highest point patterns. Two distinctive patterns created by tracking the highest point on the tongue contours can be observed in TDC1’s syllables with clustered onsets (Figures 4.46 and 4.44) and three patterns in the articulation of “lay” (Figure 4.46). Similarly, two distinctive patterns can be seen in the production of the other two selected TDC speakers (TDC9’s “pay”, “say”, “slay” and “splay” and TDC10’s “pay”) shown in Appendix III (Figures 7, 9 and 10, respectively).

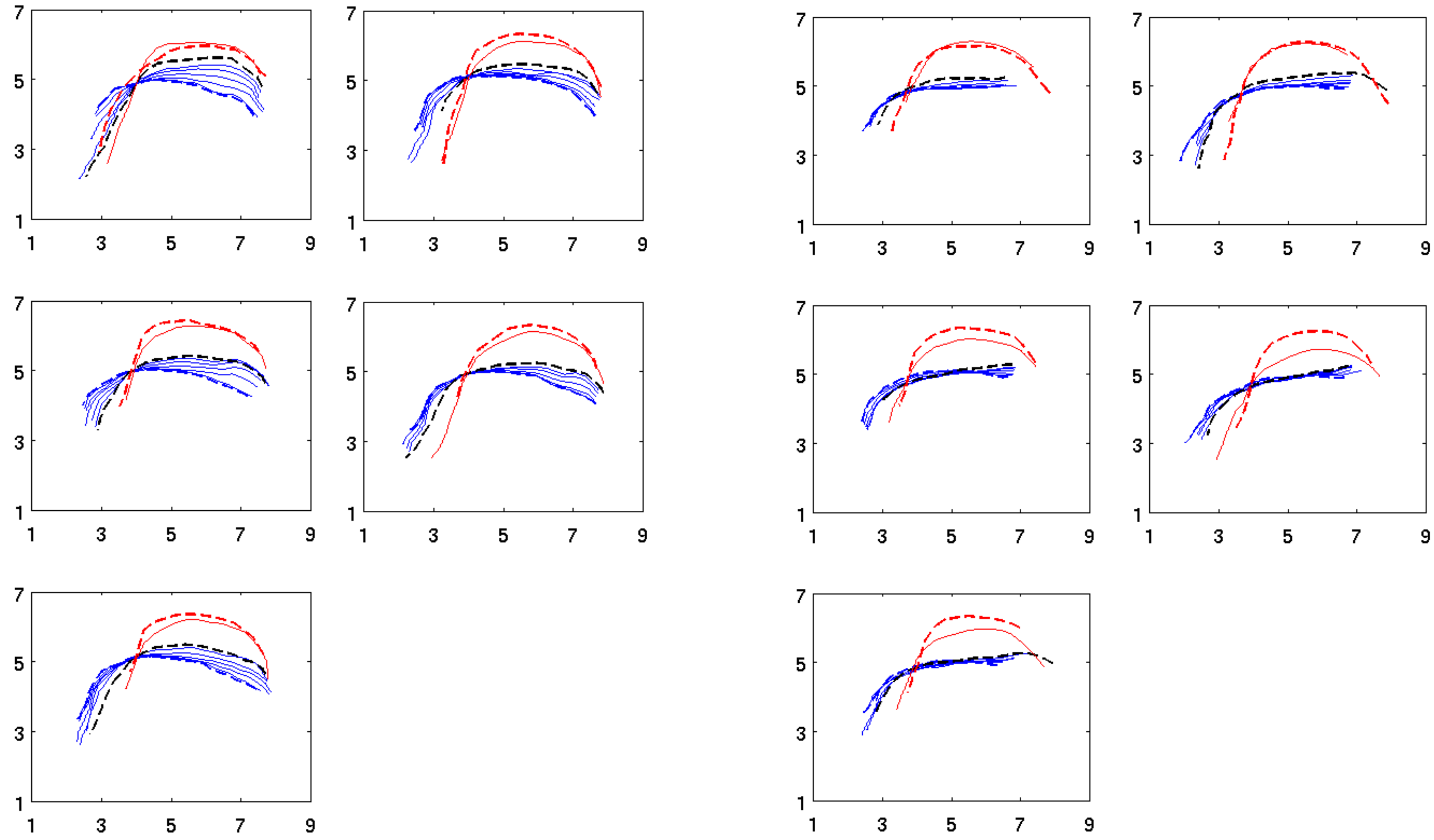


Figure 4.15: Speaker AD3's midsagittal tongue contours over "pay" (left) and "say" (right). Scale is in cm.

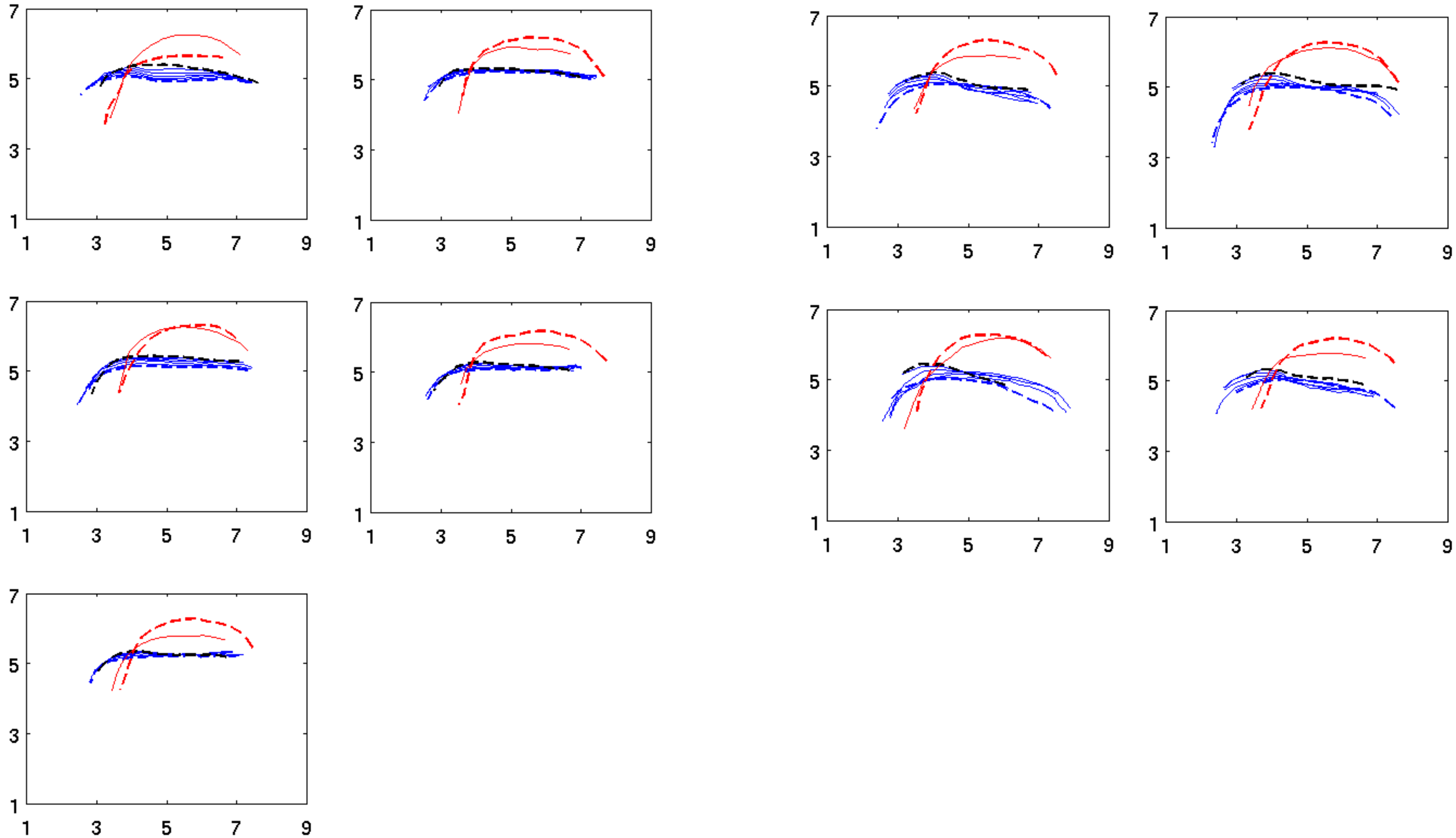


Figure 4.16: Speaker AD3's midsagittal tongue contours over “lay” (left) and “play” (right). Scale is in cm.

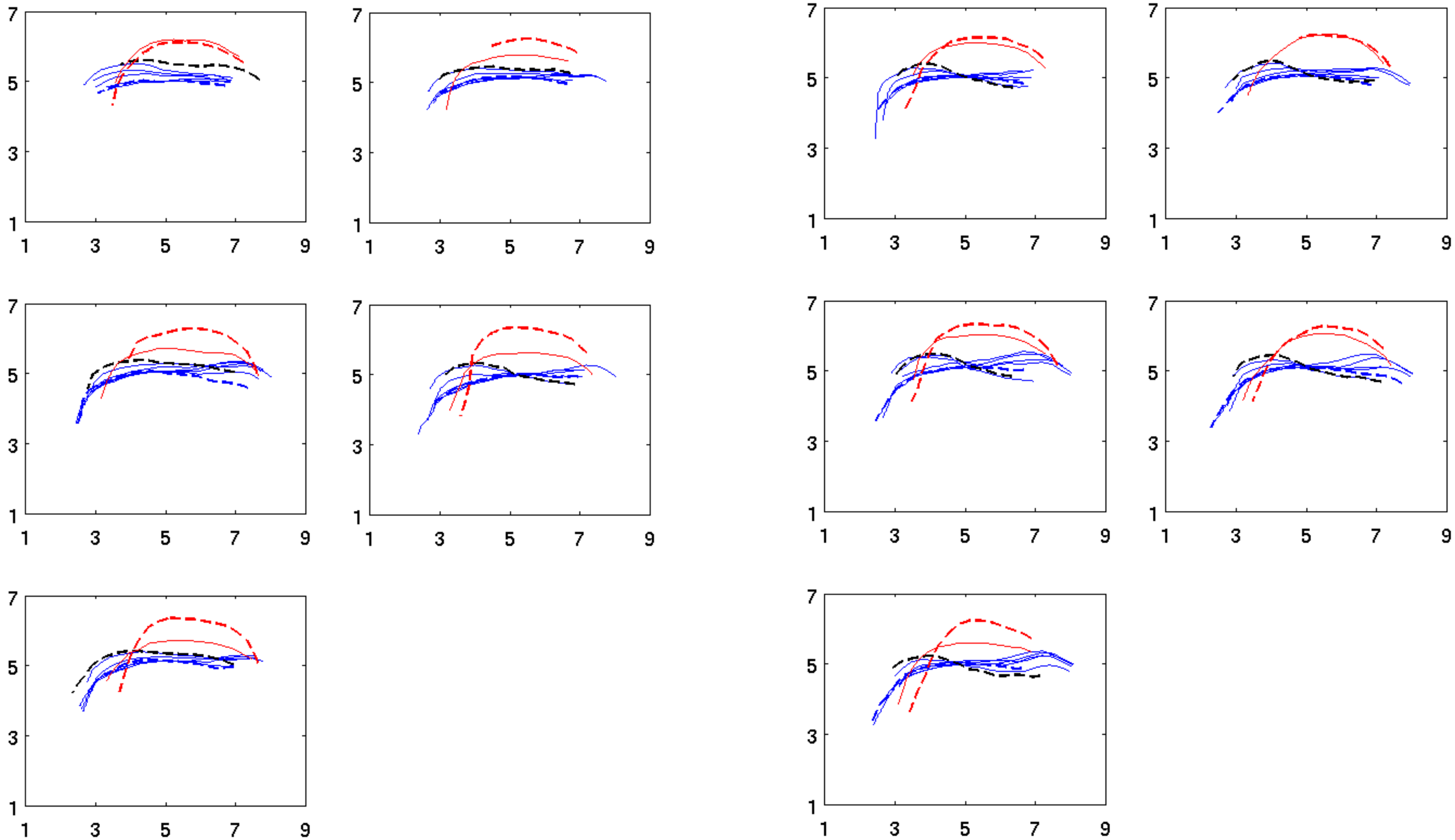


Figure 4.17: Speaker AD3's midsagittal tongue contours over “slay” (left) and “splay” (right). Scale is in cm.

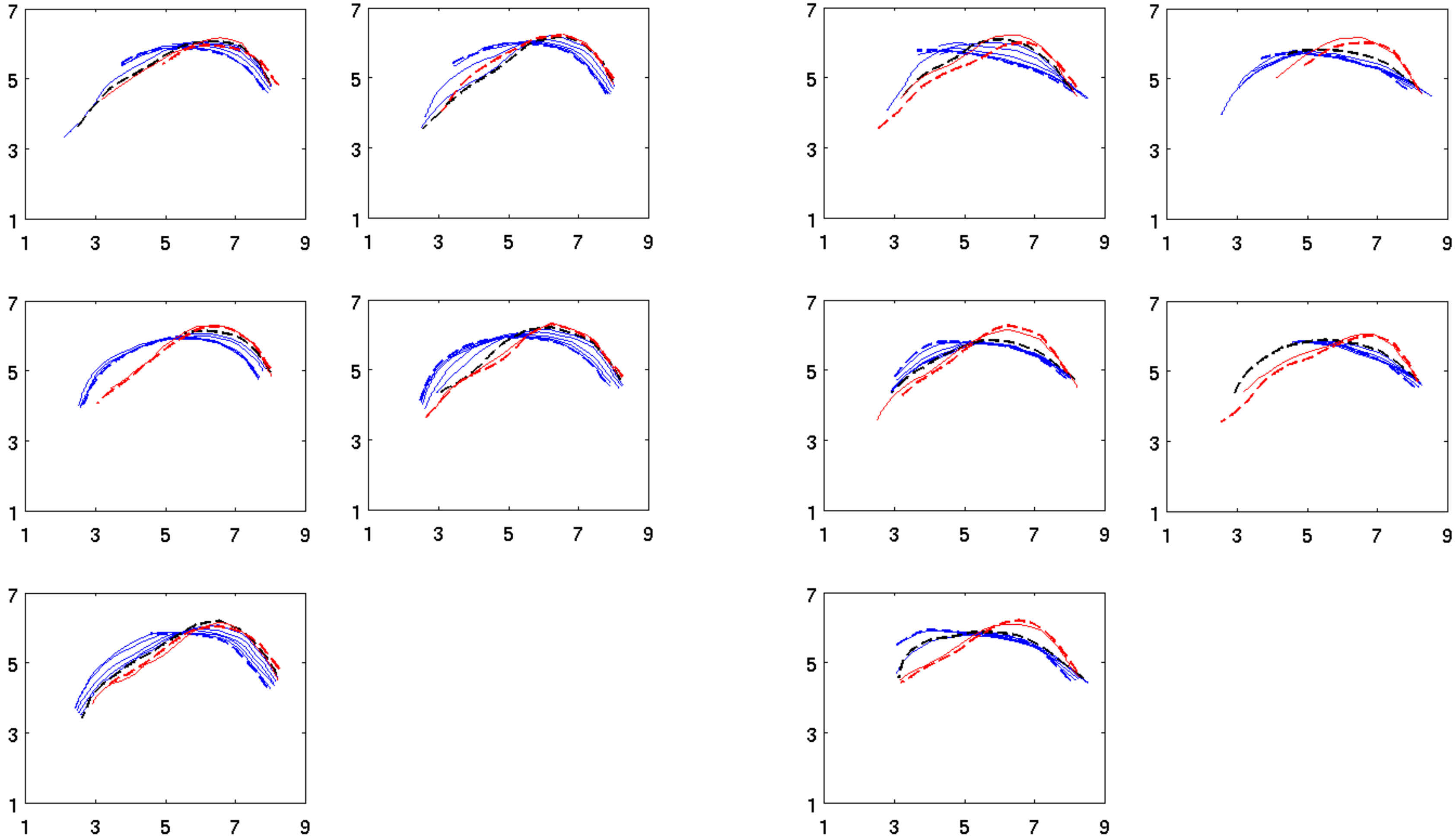


Figure 4.18: Speaker AD5's midsagittal tongue contours over "pay" (left) and "say" (right). Scale is in cm.

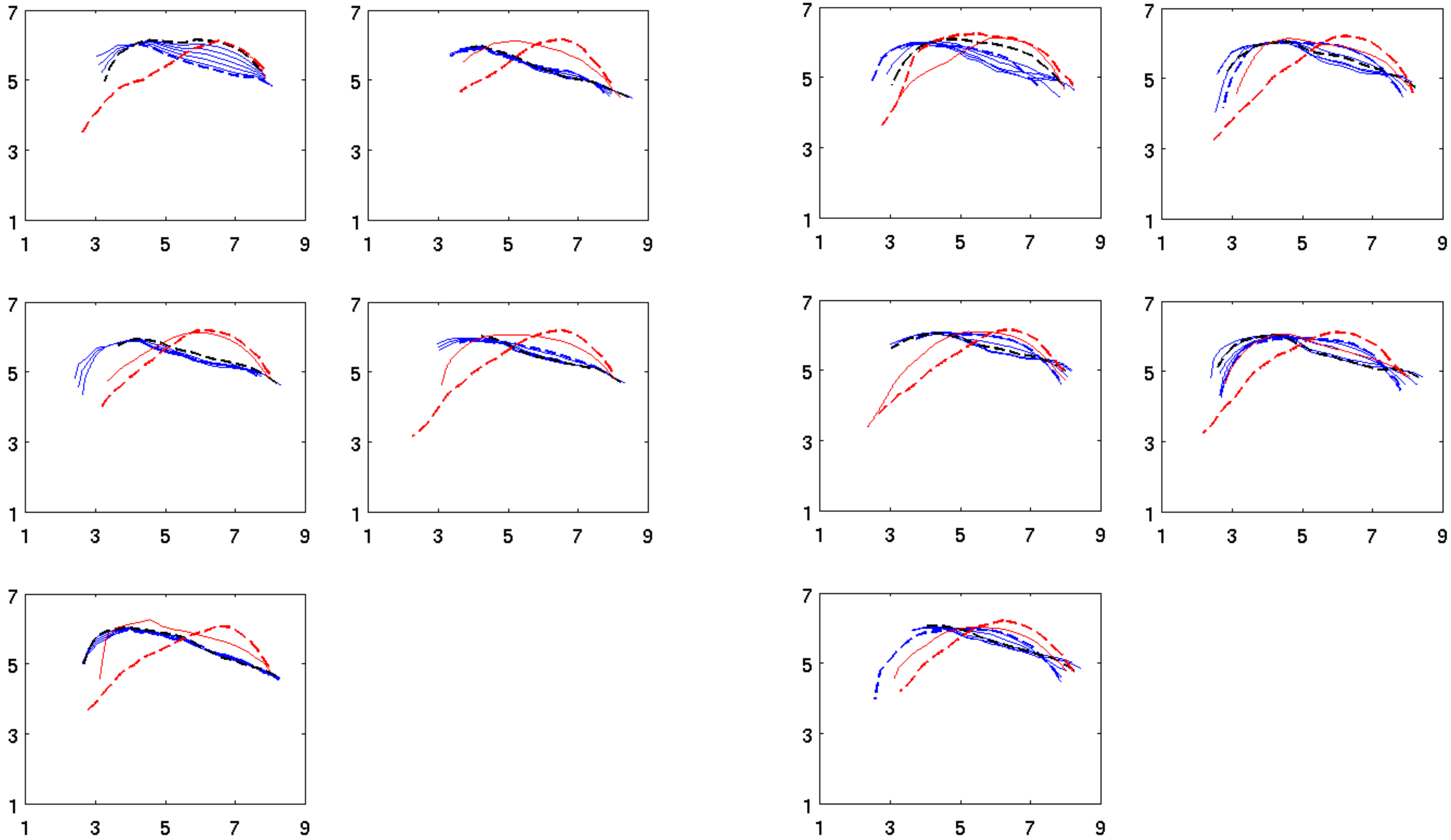


Figure 4.19: Speaker AD5's midsagittal tongue contours over "lay" (left) and "play" (right). Scale is in cm.

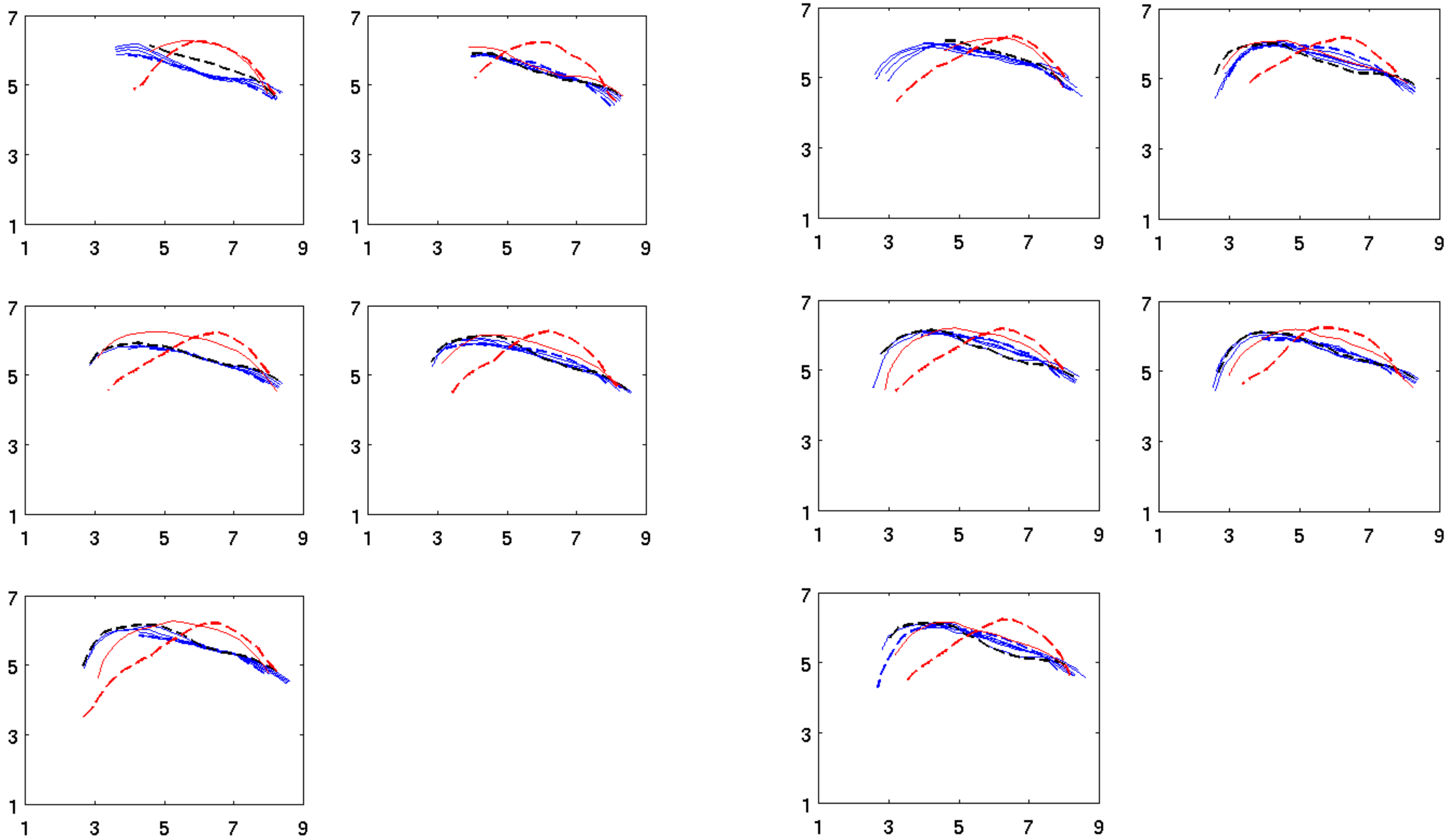


Figure 4.20: Speaker AD5's midsagittal tongue contours over “slay” (left) and “splay” (right). Scale is in cm.

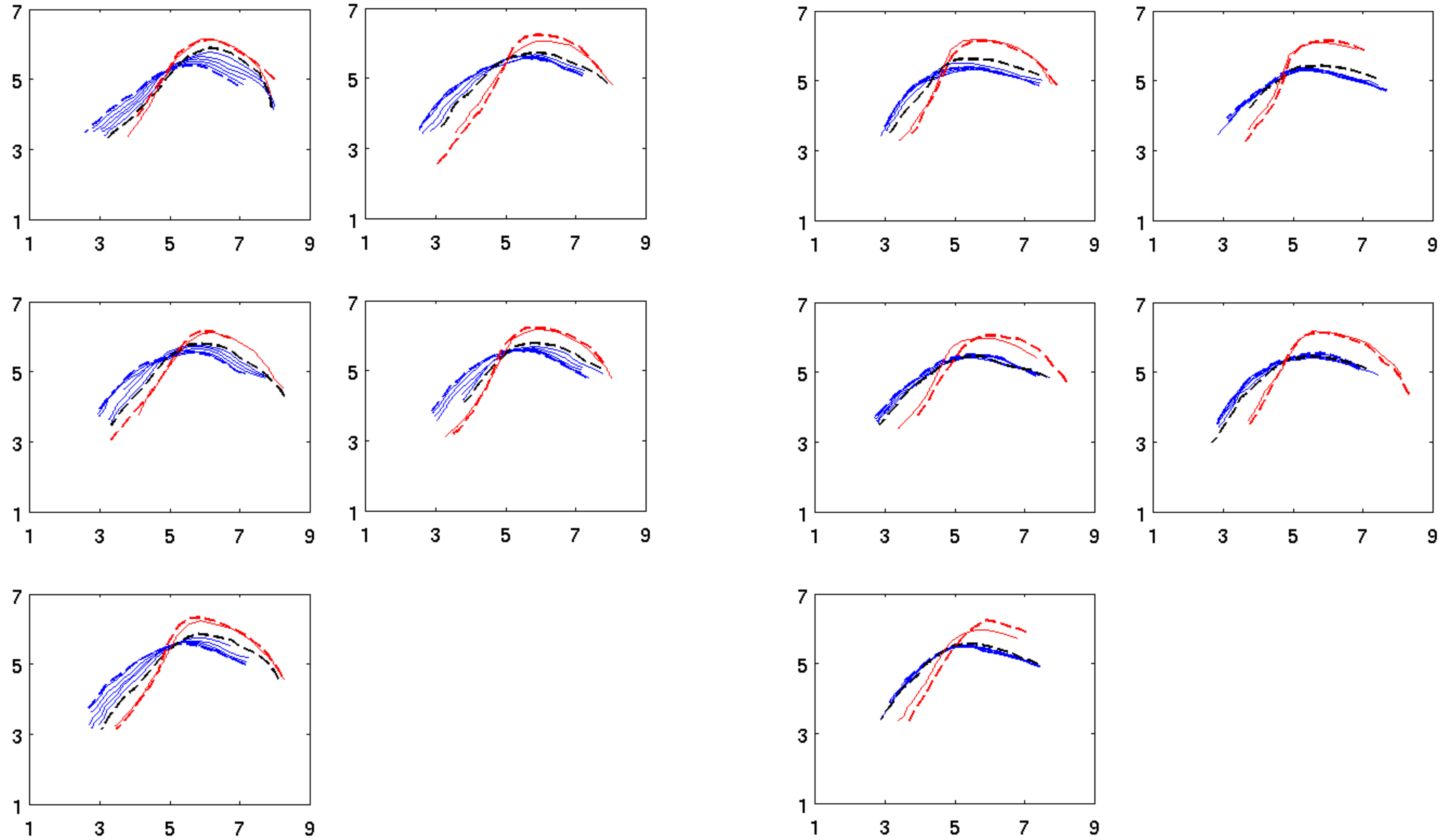


Figure 4.21: Speaker AD9's midsagittal tongue contours over "pay" (left) and "say" (right). Scale is in cm.

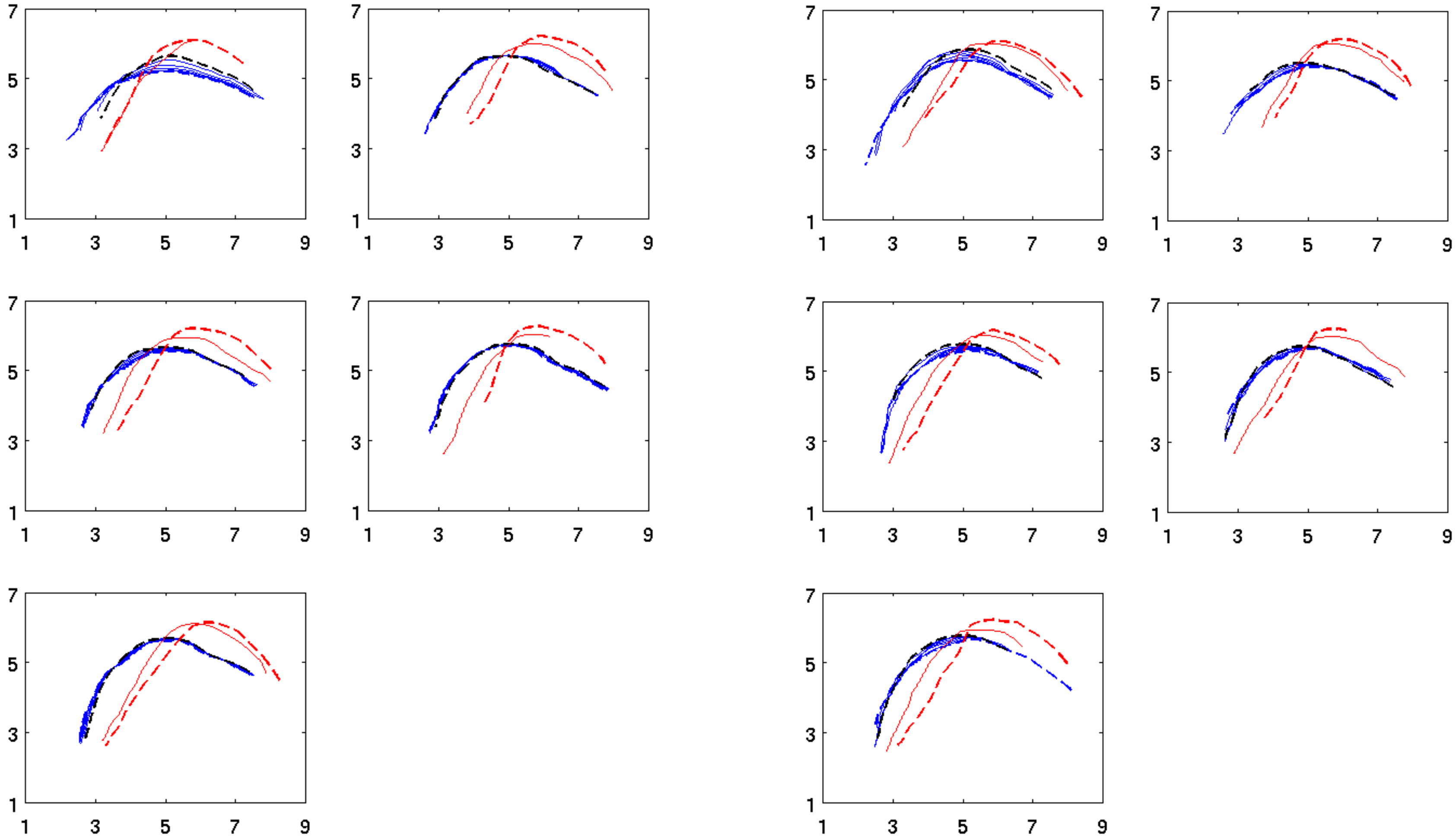


Figure 4.22: Speaker AD9's midsagittal tongue contours over "lay" (left) and "play" (right). Scale is in cm.

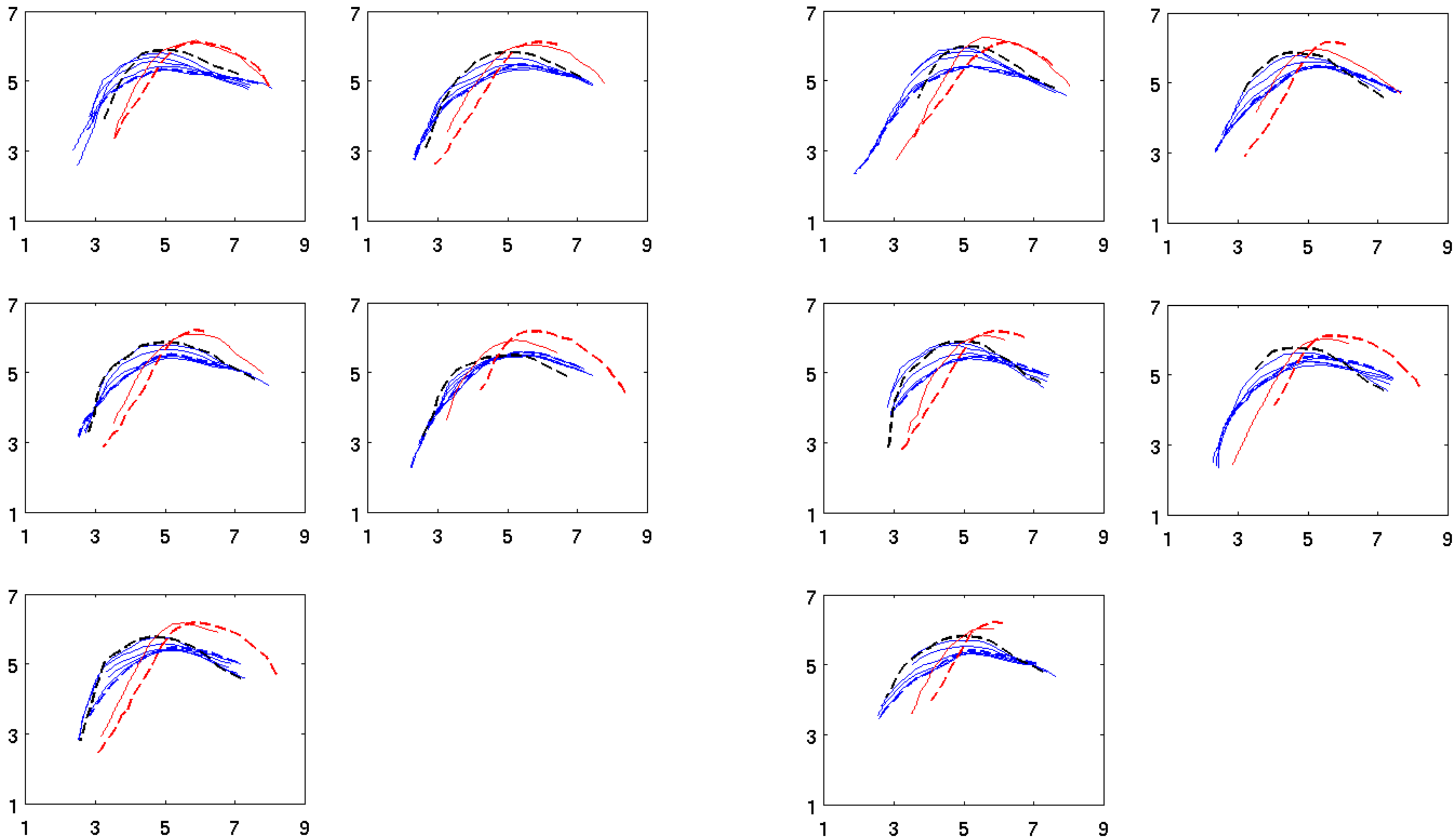


Figure 4.23: Speaker AD9's midsagittal tongue contours over “slay” (left) and “splay” (right). Scale is in cm.

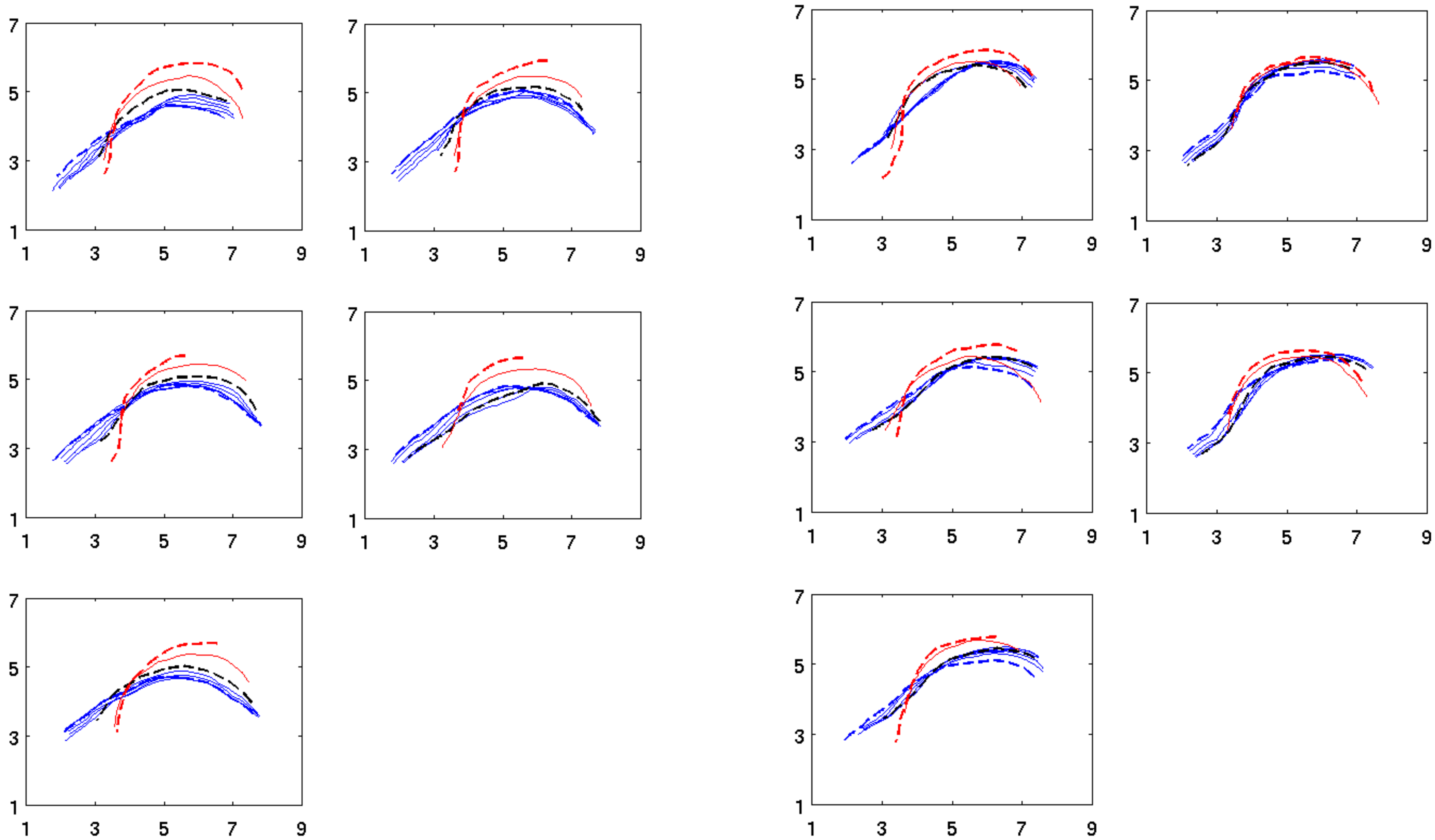


Figure 4.24: Speaker TDC1's midsagittal tongue contours over “pay” (left) and “say” (right). Scale is in cm.

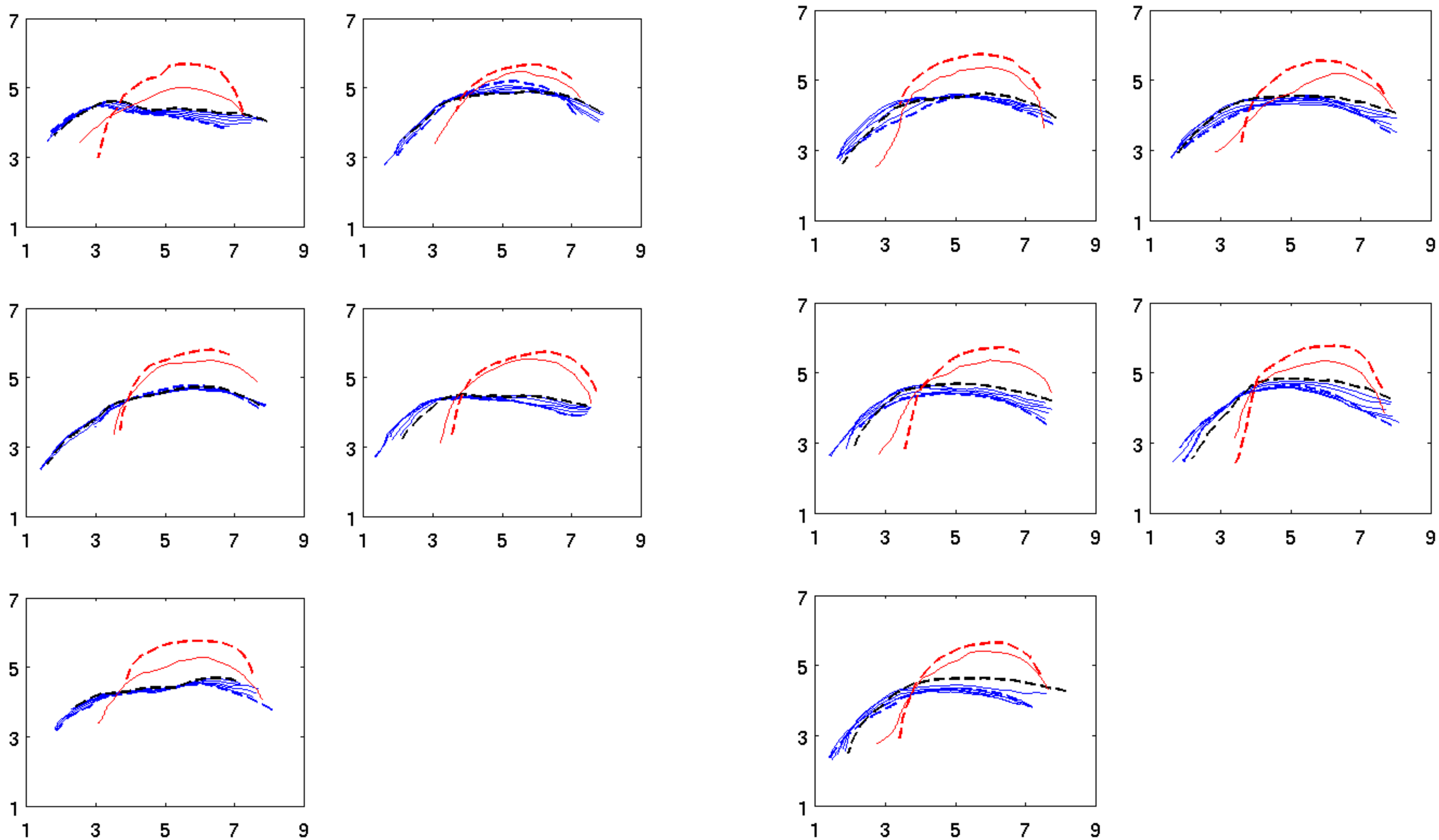


Figure 4.25: Speaker TDC1’s midsagittal tongue contours over “lay” (left) and “play” (right). Scale is in cm.

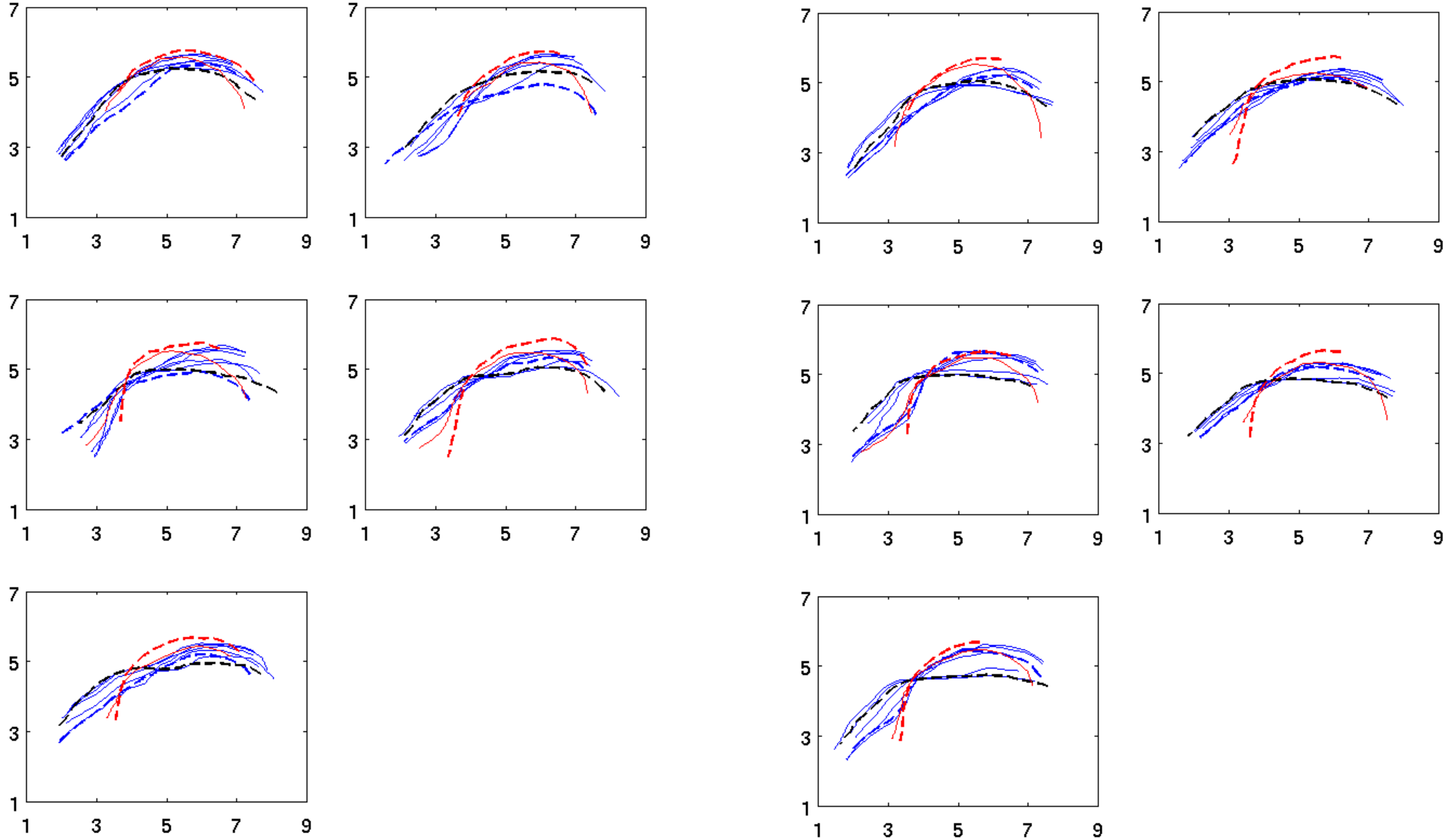


Figure 4.26: Speaker TDC1's midsagittal tongue contours over “slay” (left) and “splay” (right). Scale is in cm.

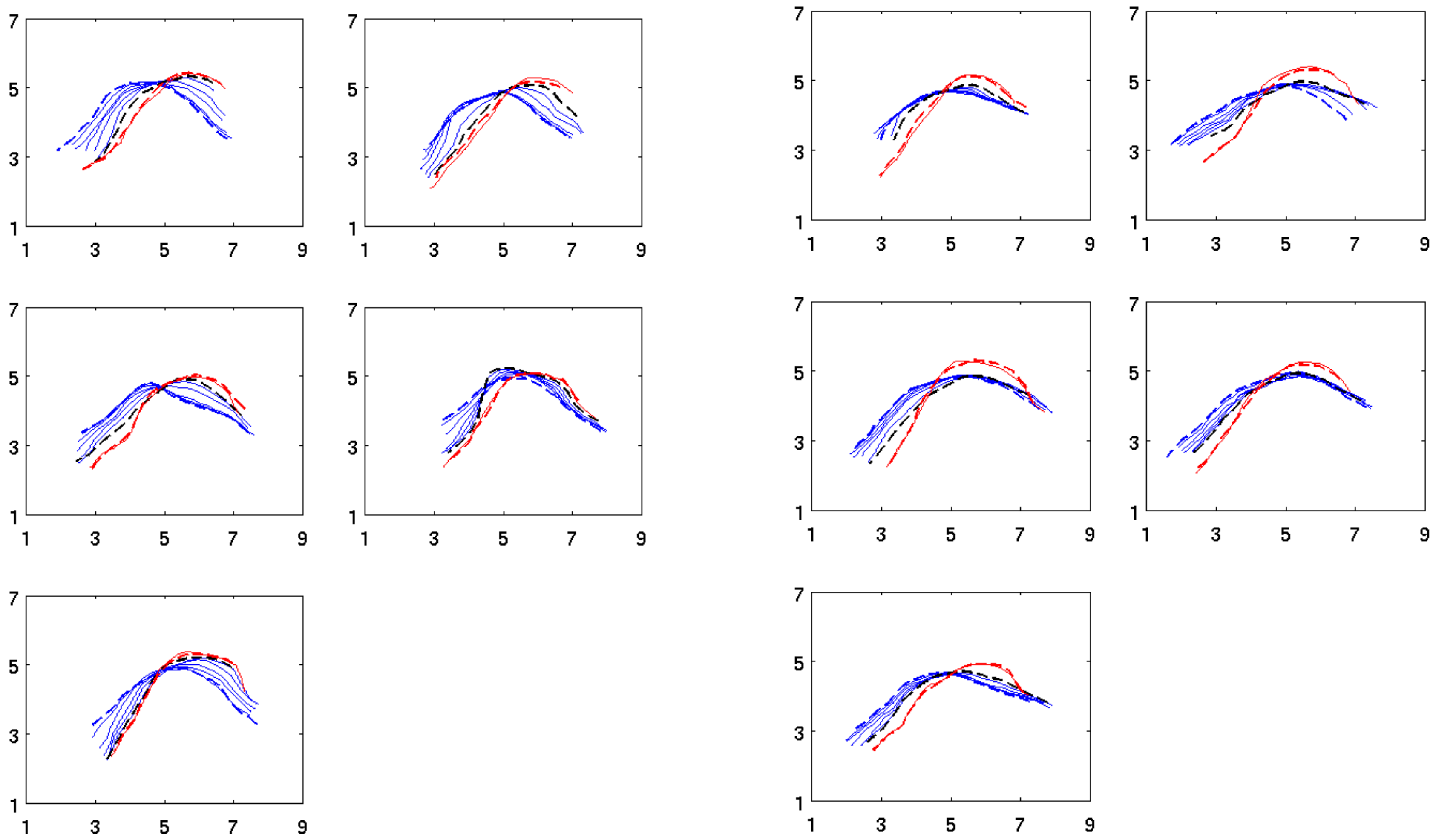


Figure 4.27: Speaker TDC9's midsagittal tongue contours over "pay" (left) and "say" (right). Scale is in cm.

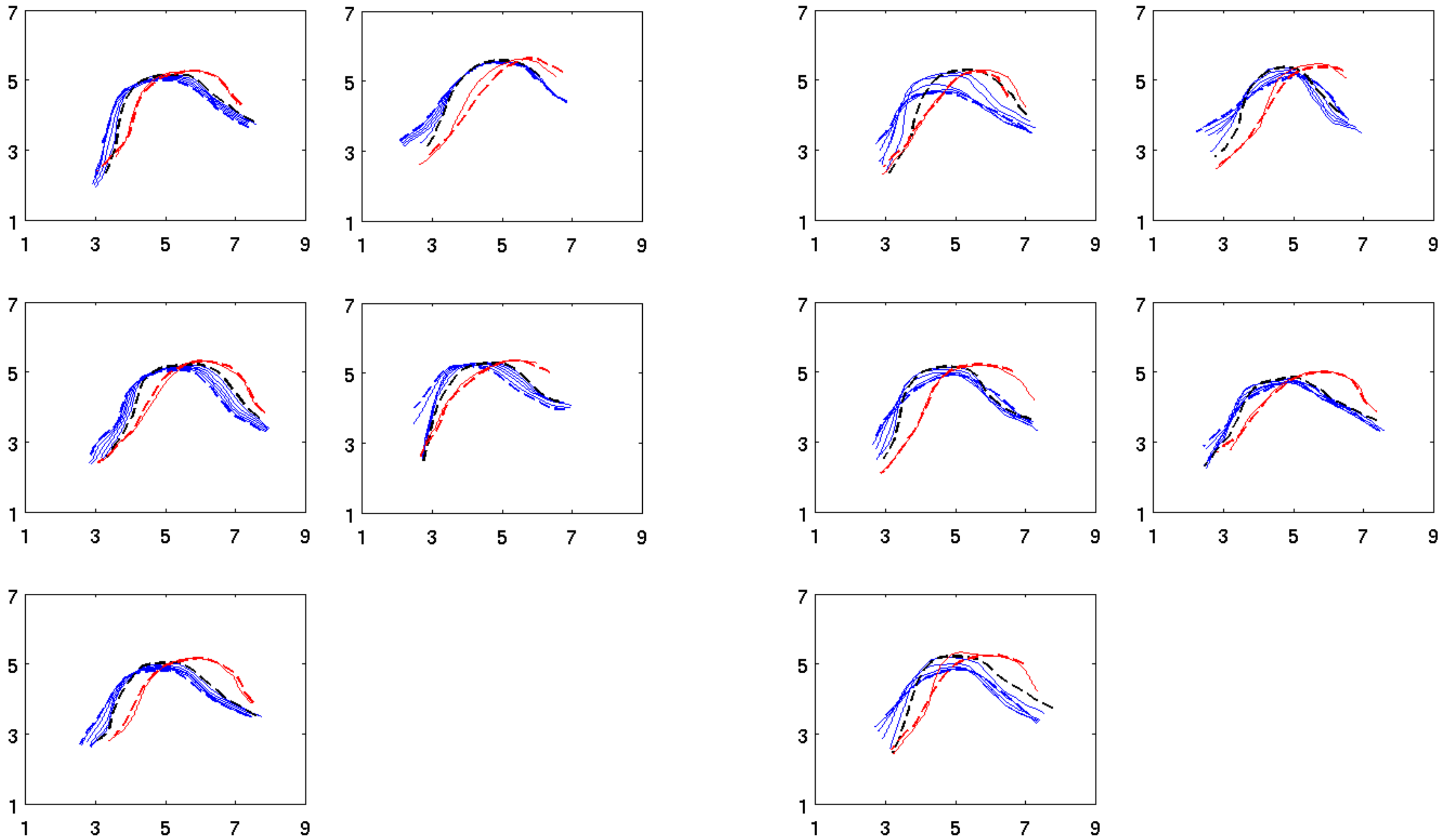


Figure 4.28: Speaker TDC9’s midsagittal tongue contours over “lay” (left) and “play” (right). Scale is in cm.

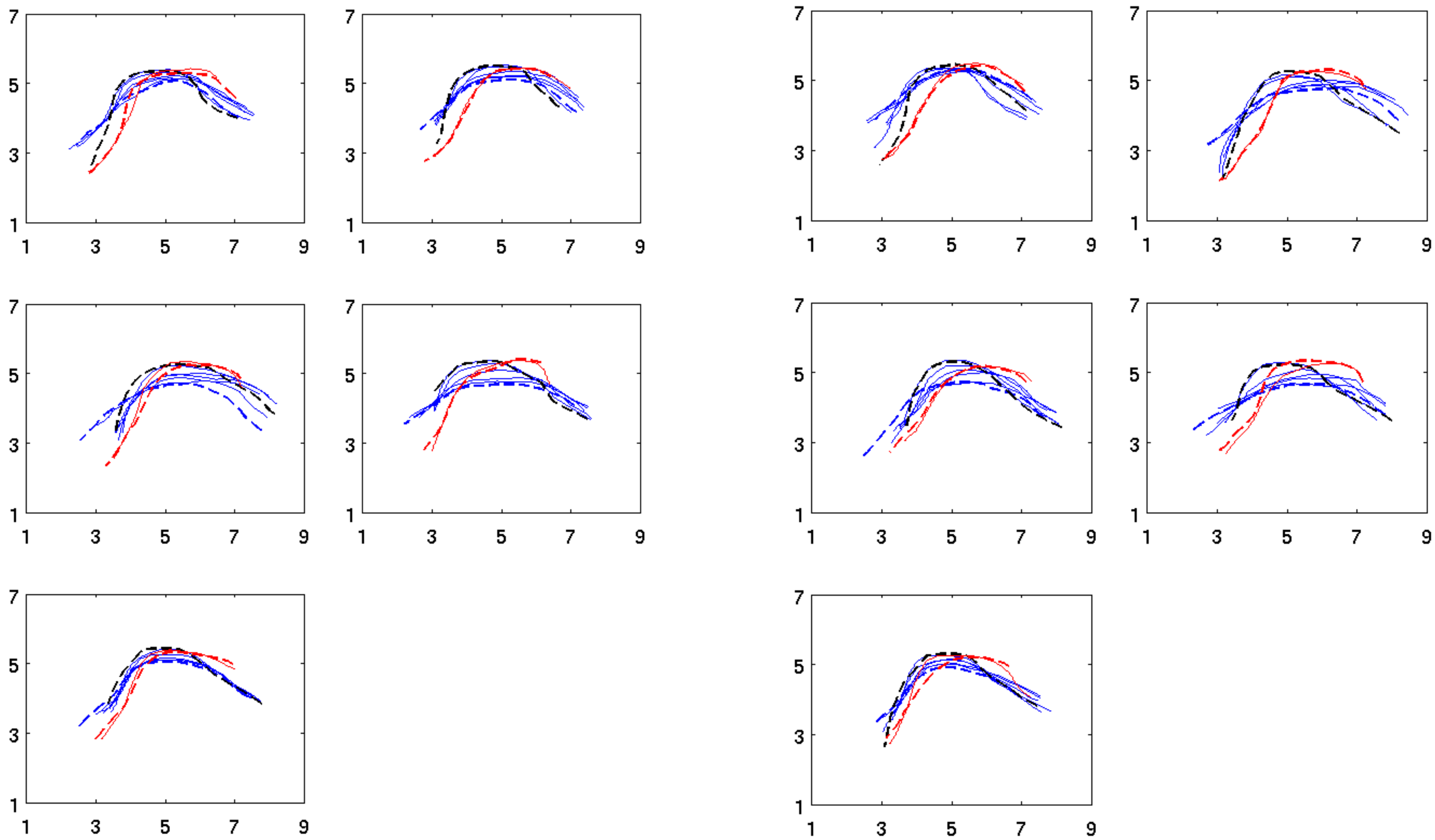


Figure 4.29: Speaker TDC9's midsagittal tongue contours over "slay" (left) and "splay" (right). Scale is in cm.

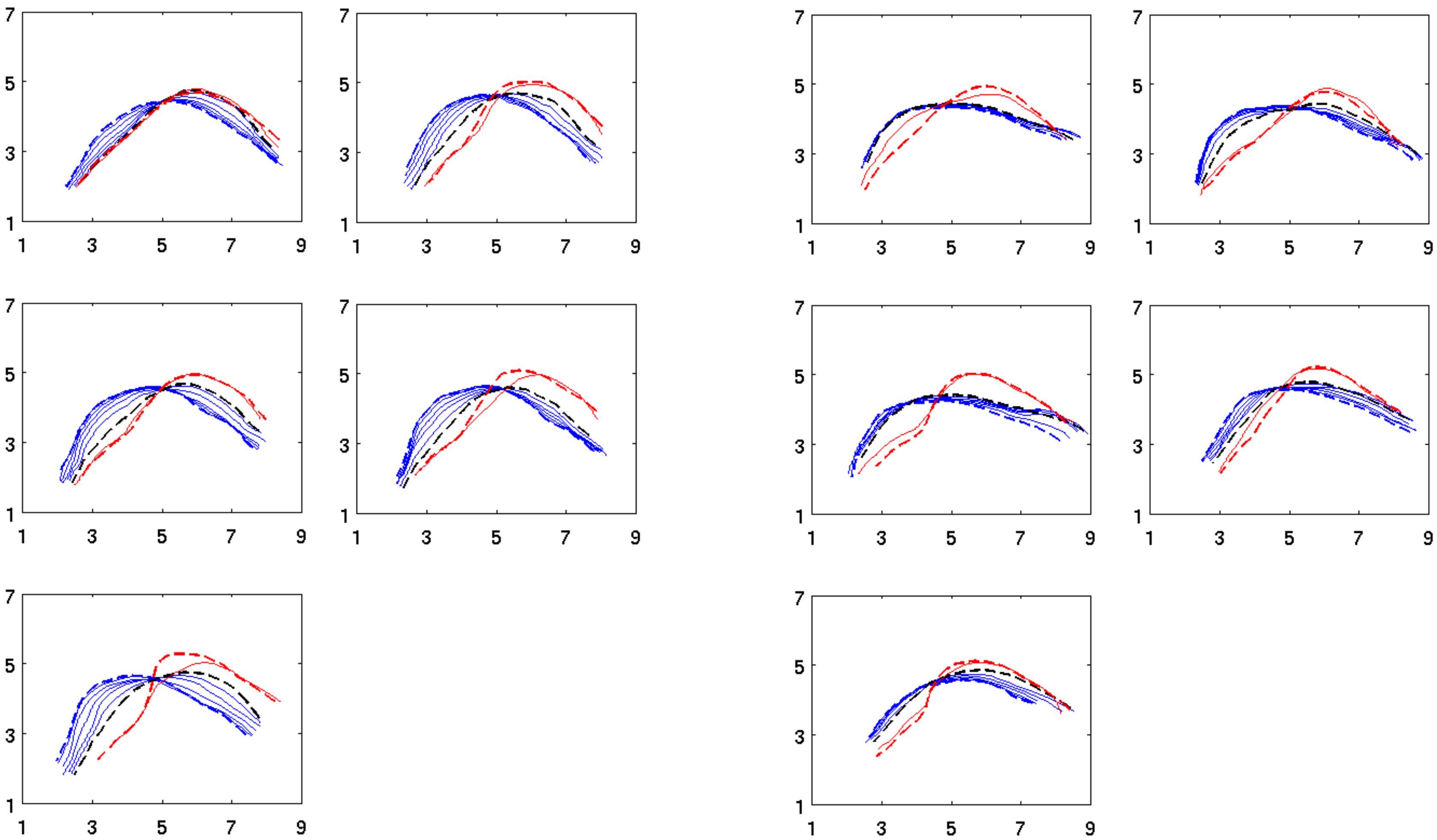


Figure 4.30: Speaker TDC10's midsagittal tongue contours over “pay” (left) and “say” (right). Scale is in cm.

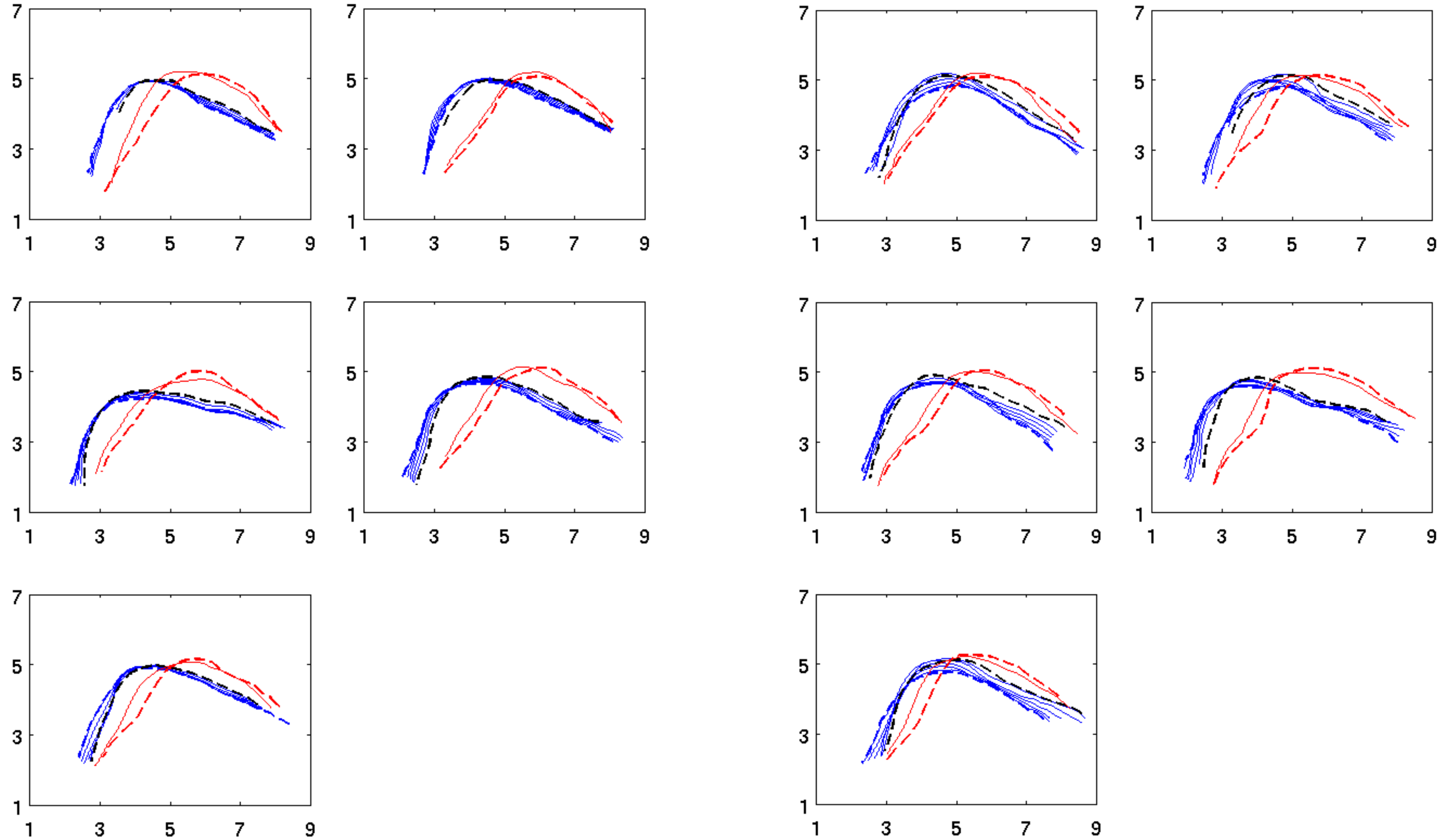


Figure 4.31: Speaker TDC10's midsagittal tongue contours over “lay” (left) and “play” (right). Scale is in cm.

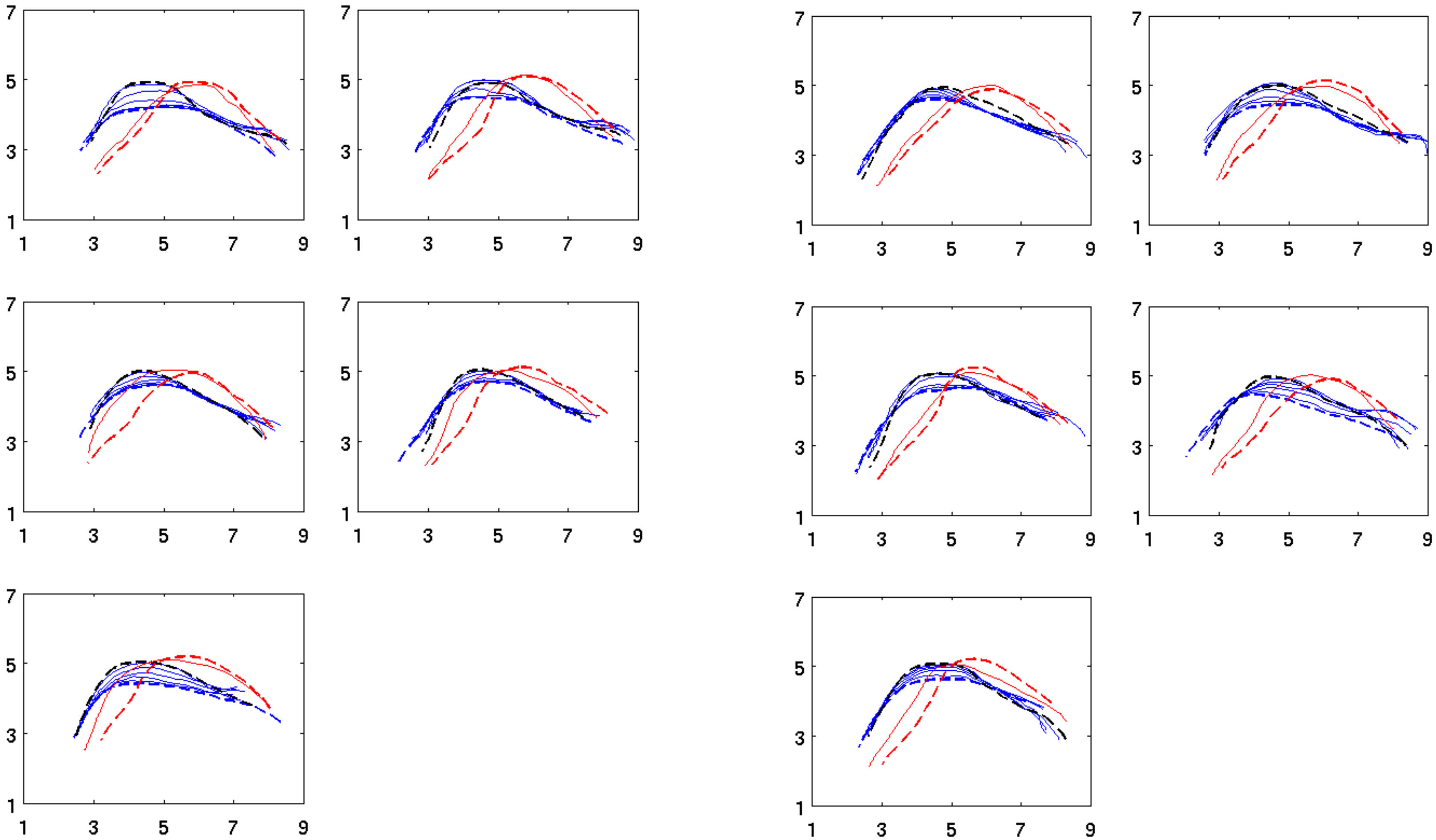


Figure 4.32: Speaker TDC10’s midsagittal tongue contours over “slay” (left) and “splay” (right). Scale is in cm.

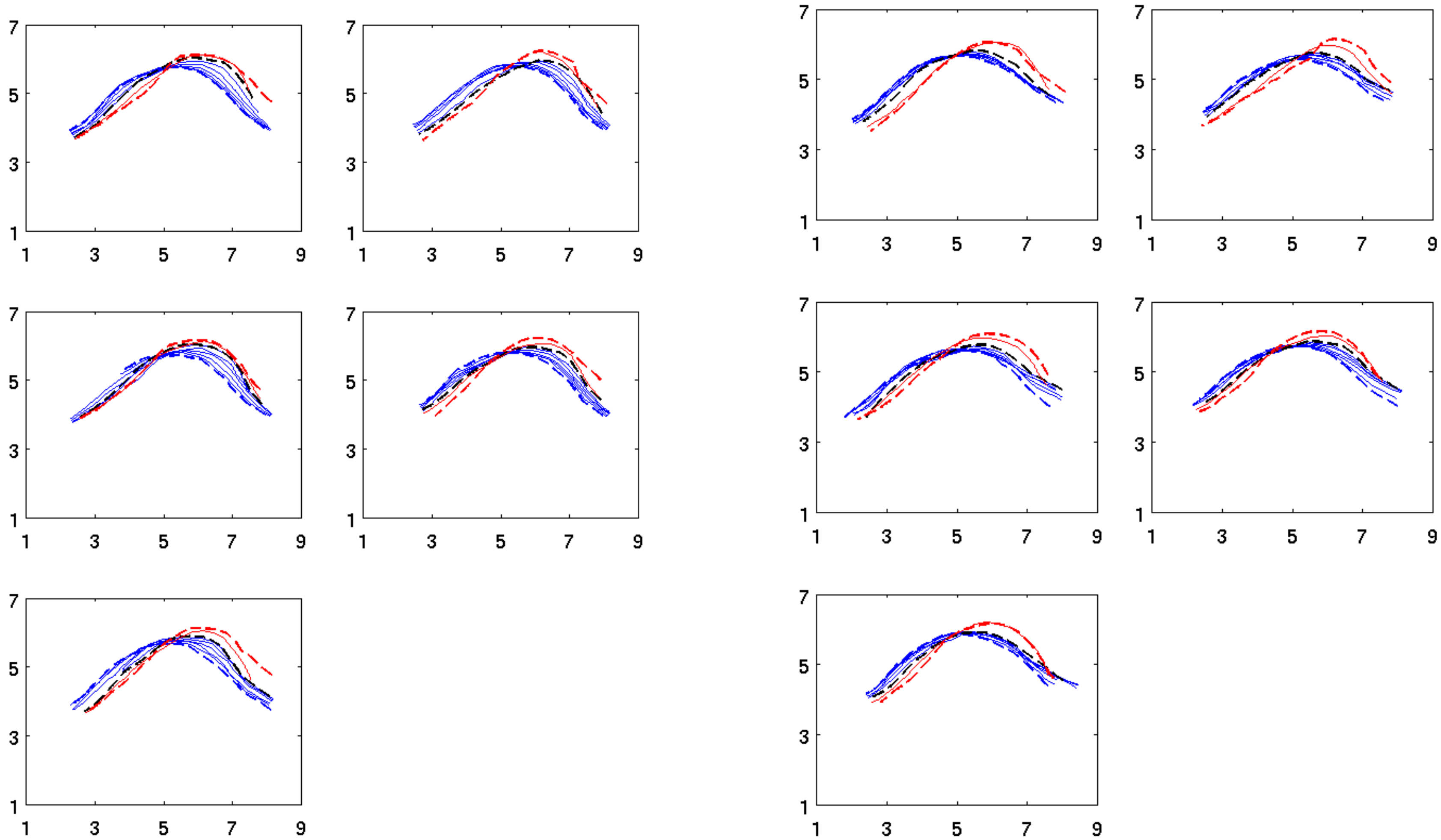


Figure 4.33: Speaker CAS1's midsagittal tongue contours over "pay" (left) and "say" (right). Scale is in cm.

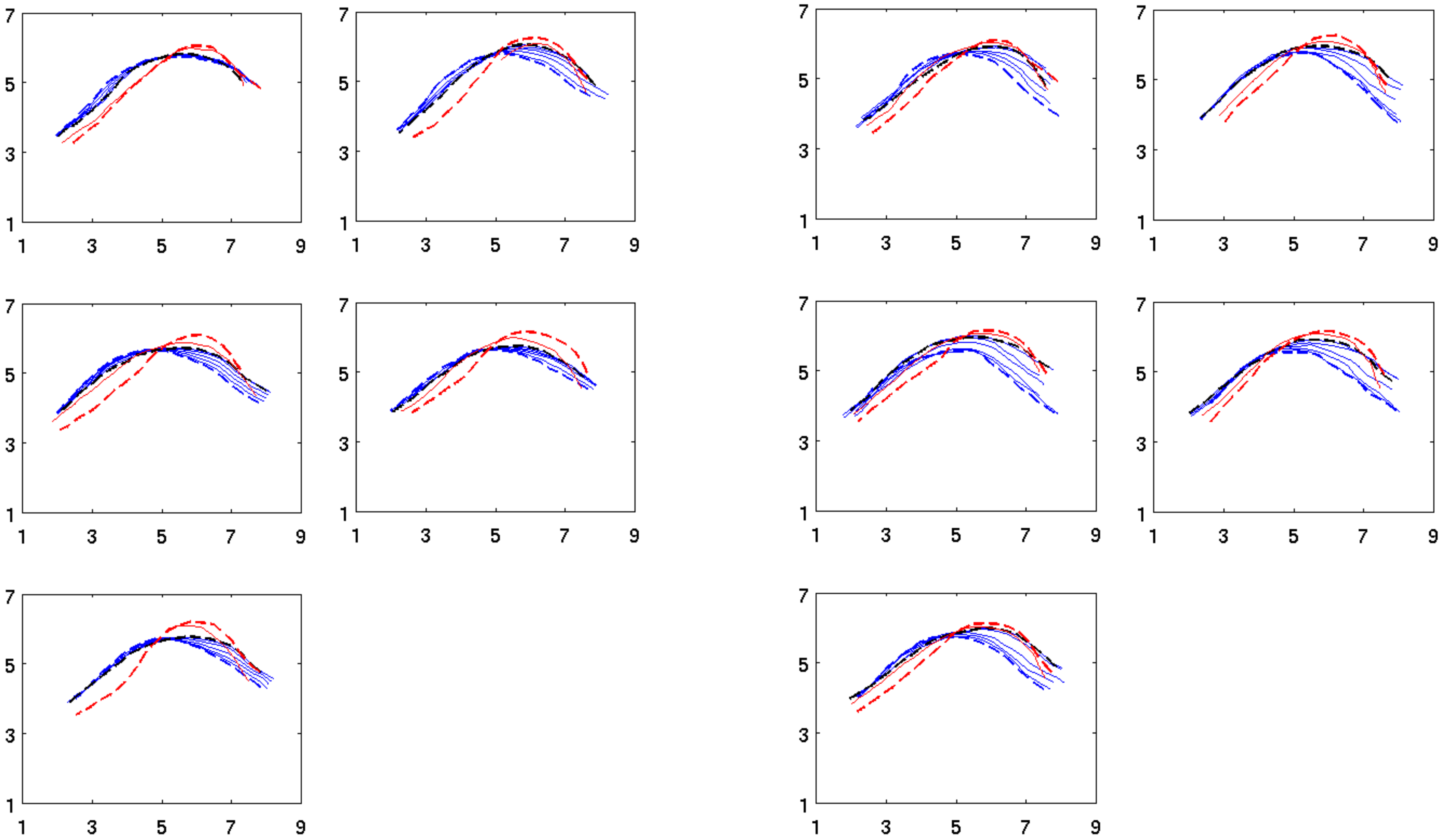


Figure 4.34: Speaker CAS1’s midsagittal tongue contours over “lay” (left) and “play” (right). Scale is in cm.

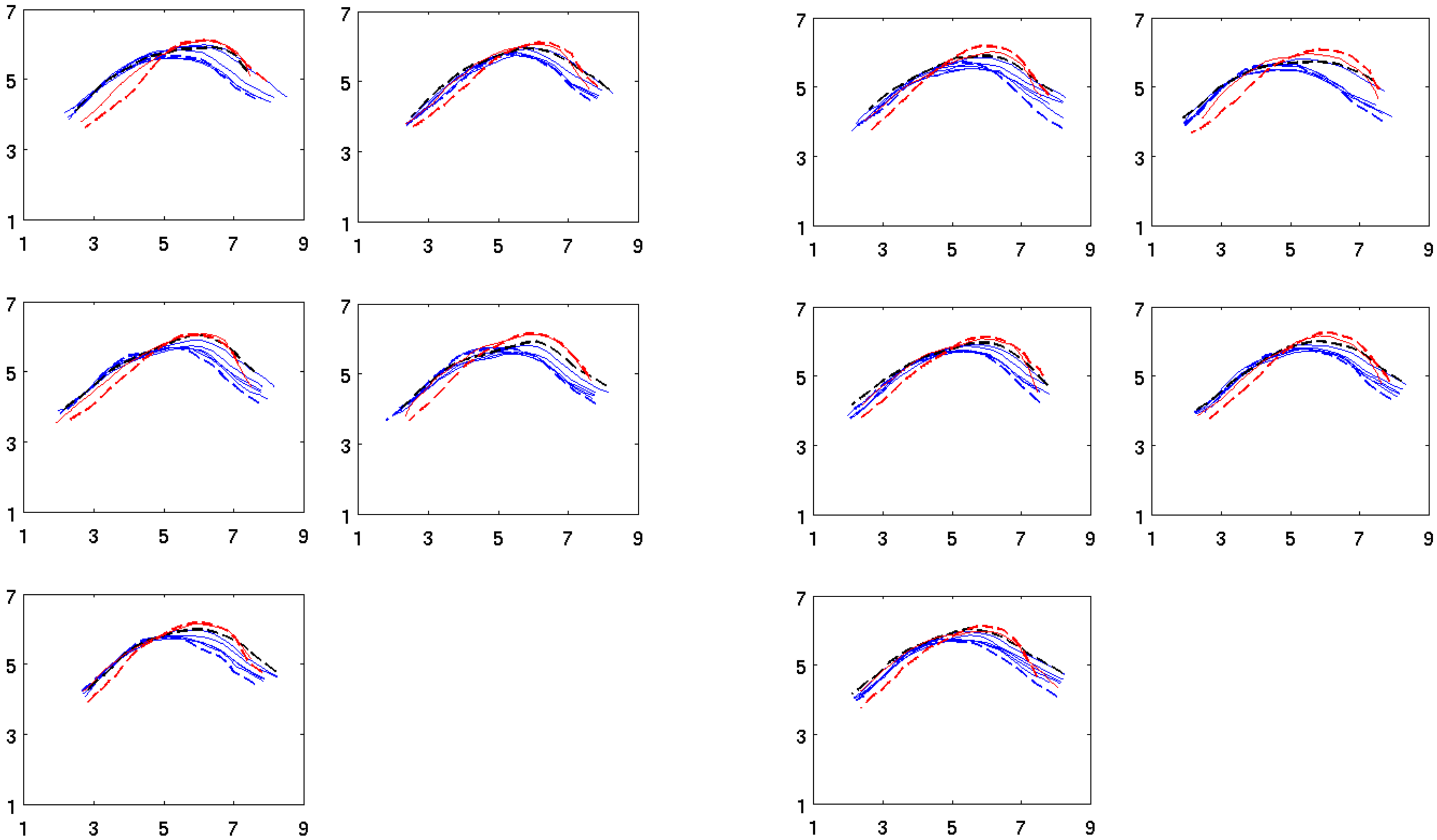


Figure 4.35: Speaker CAS1's midsagittal tongue contours over “slay” (left) and “splay” (right). Scale is in cm.

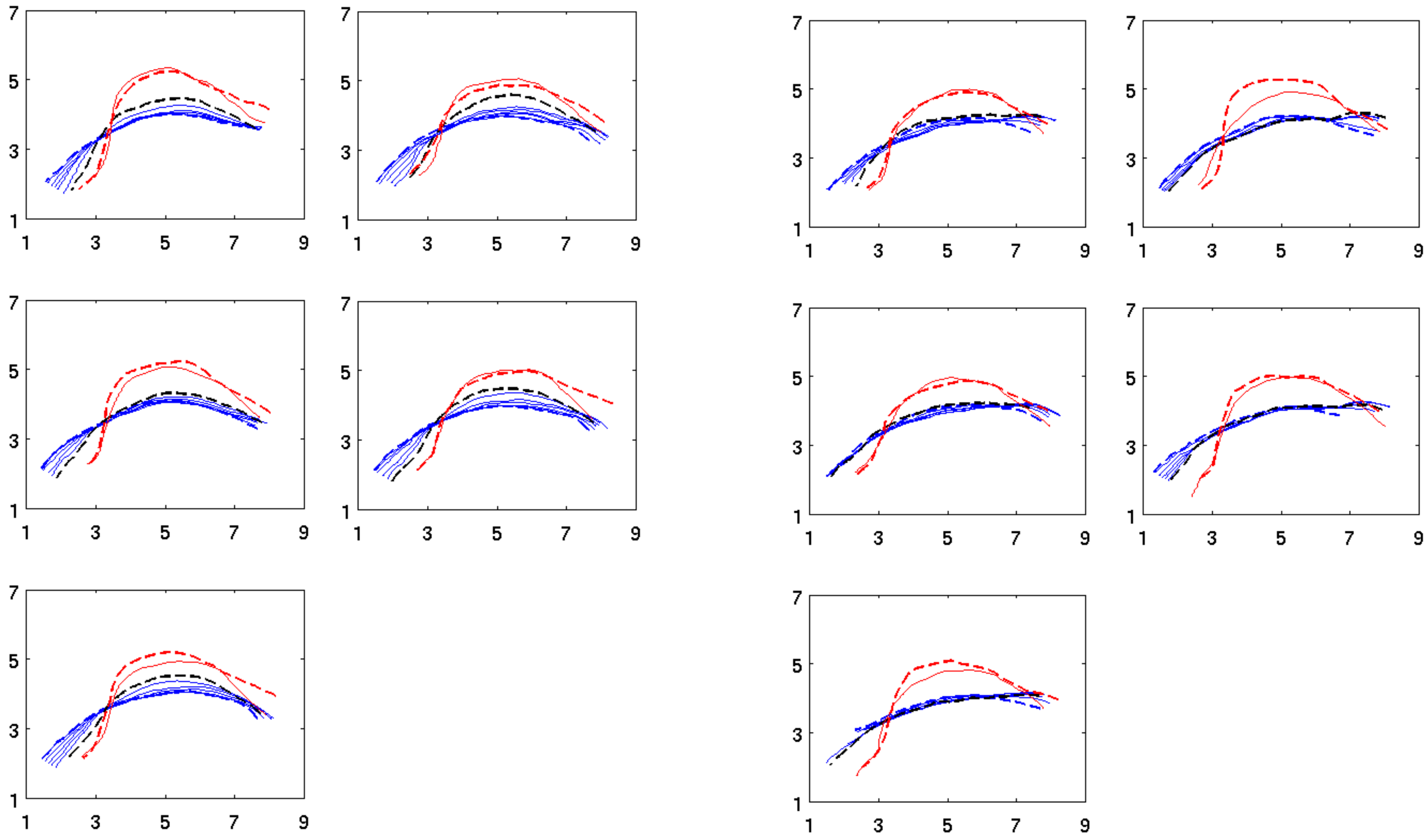


Figure 4.36: Speaker CAS2's midsagittal tongue contours over “pay” (left) and “say” (right). Scale is in cm.

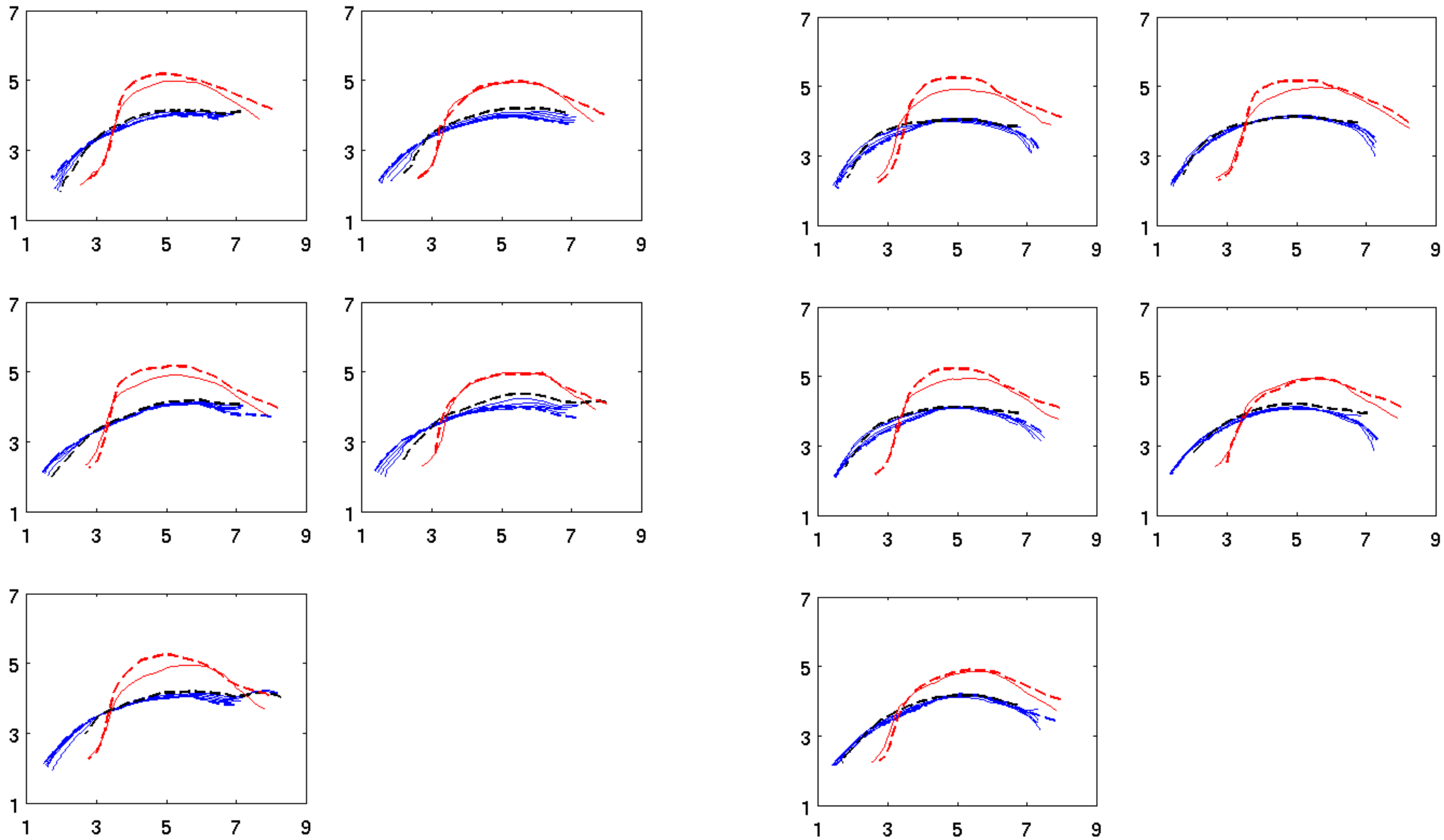


Figure 4.37: Speaker CAS2's midsagittal tongue contours over "lay" (left) and "play" (right). Scale is in cm.

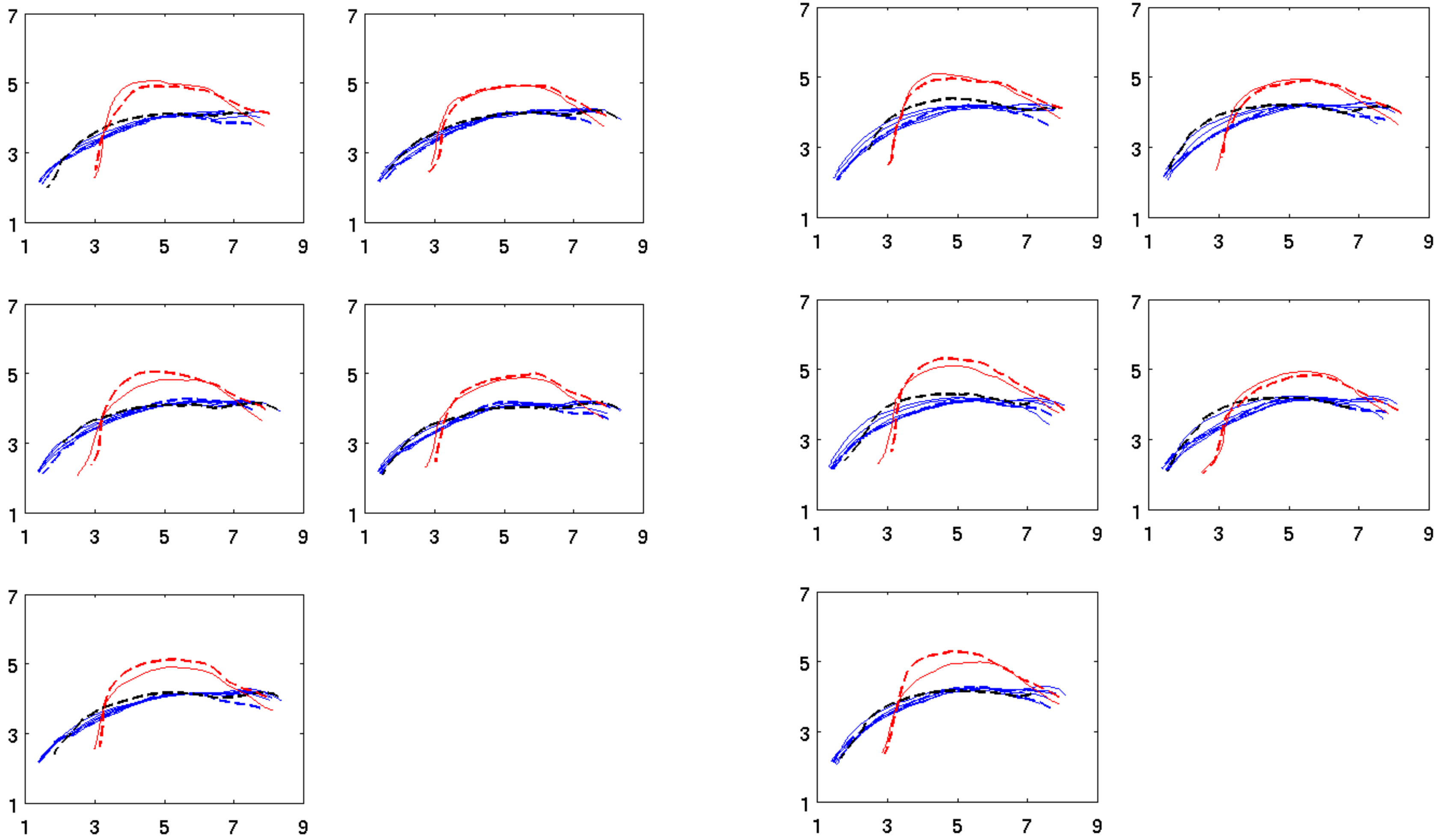


Figure 4.38: Speaker CAS2's midsagittal tongue contours over "slay" (left) and "splay" (right). Scale is in cm.

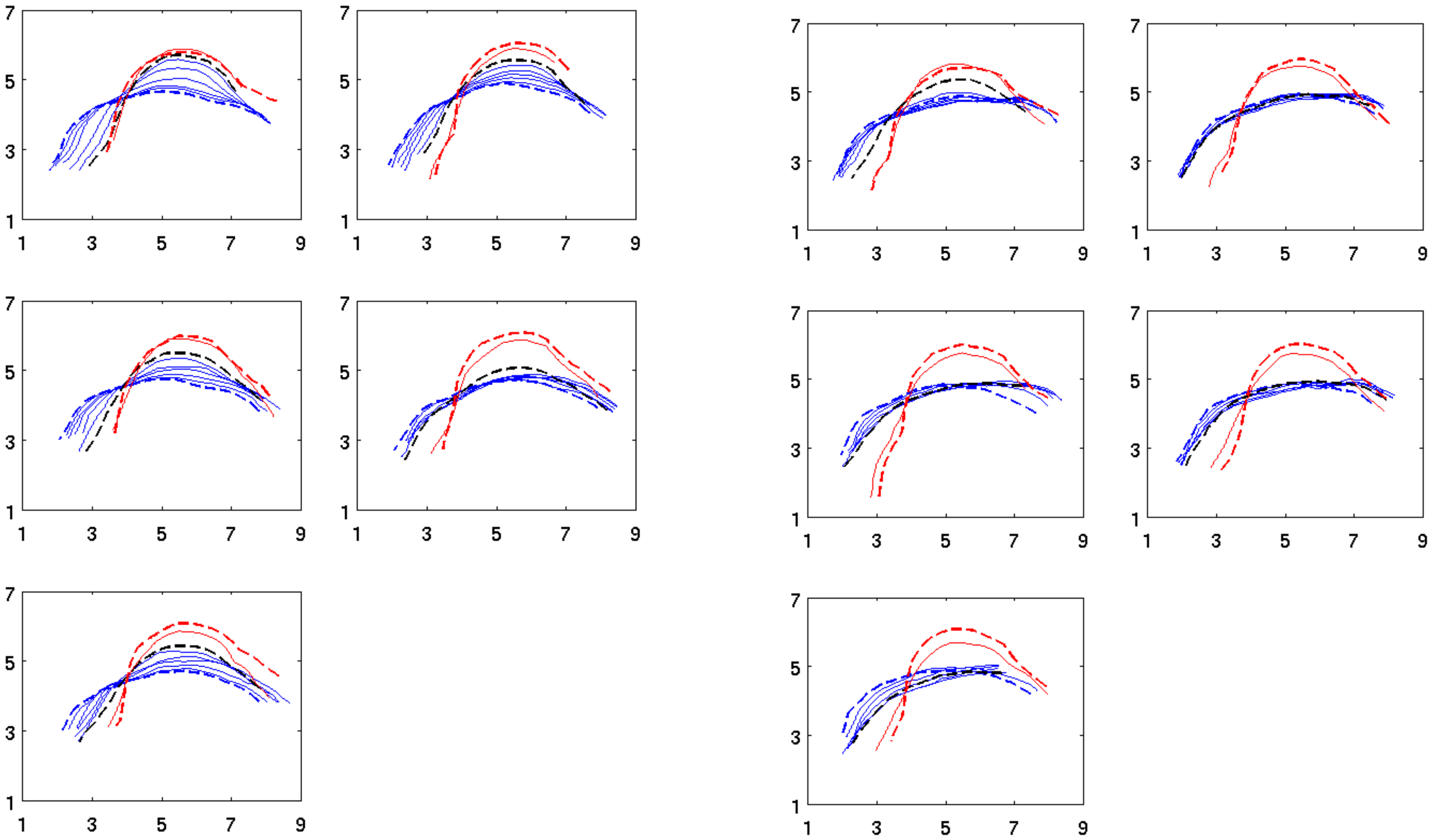


Figure 4.39: Speaker CAS3's midsagittal tongue contours over “pay” (left) and “say” (right). Scale is in cm.

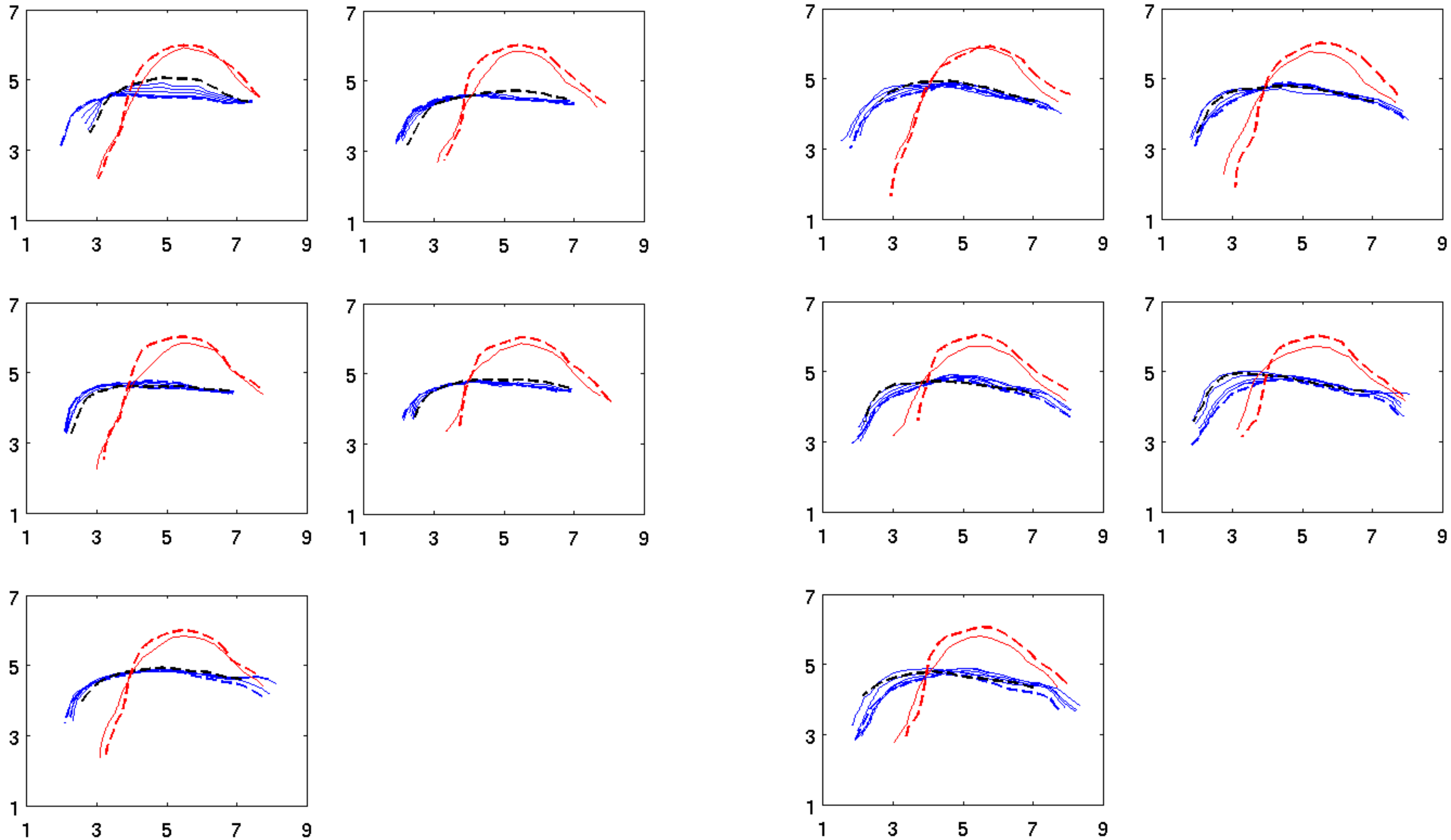


Figure 4.40: Speaker CAS3’s midsagittal tongue contours over “lay” (left) and “play” (right). Scale is in cm.

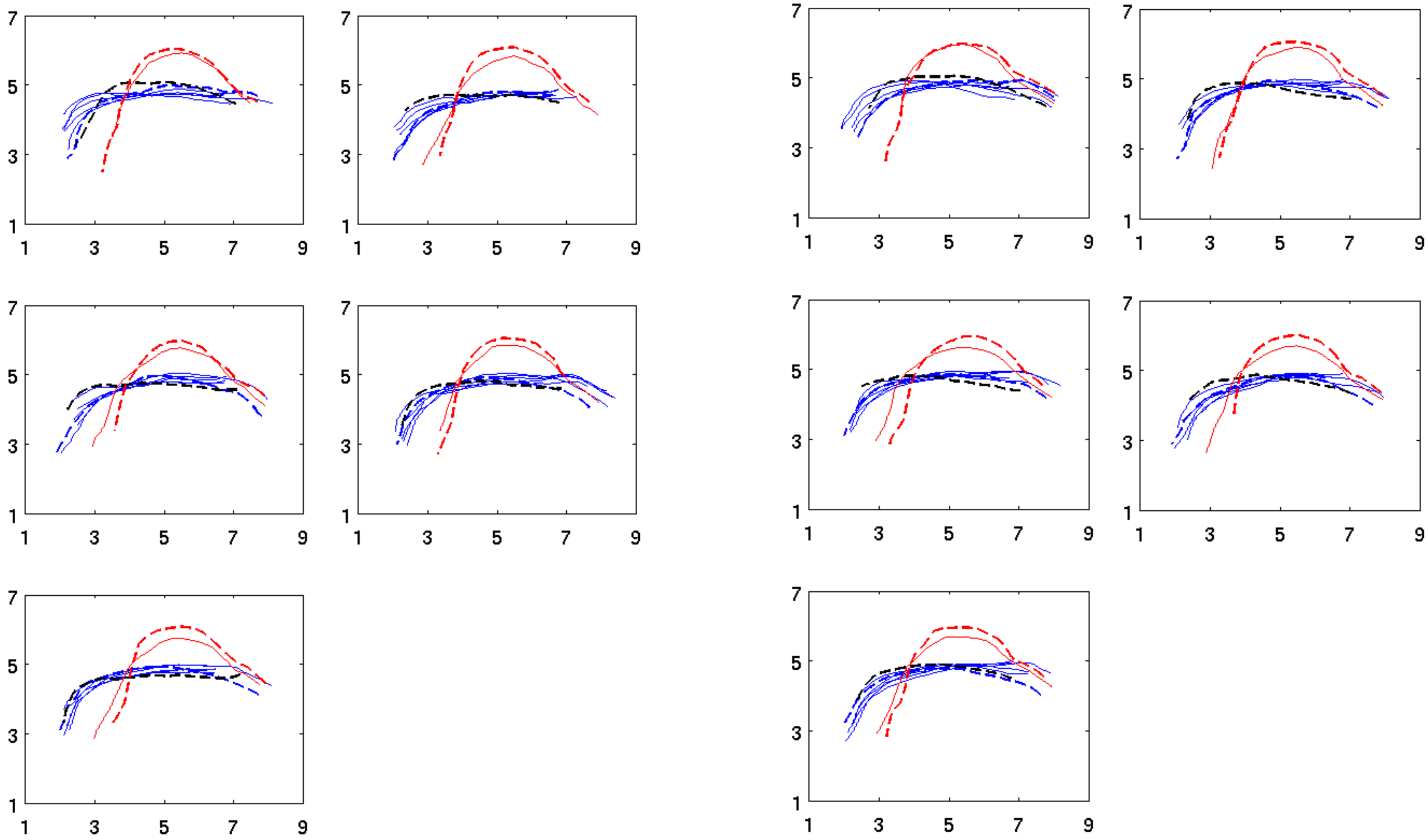


Figure 4.41: Speaker CAS3's midsagittal tongue contours over “slay” (left) and “splay” (right). Scale is in cm.

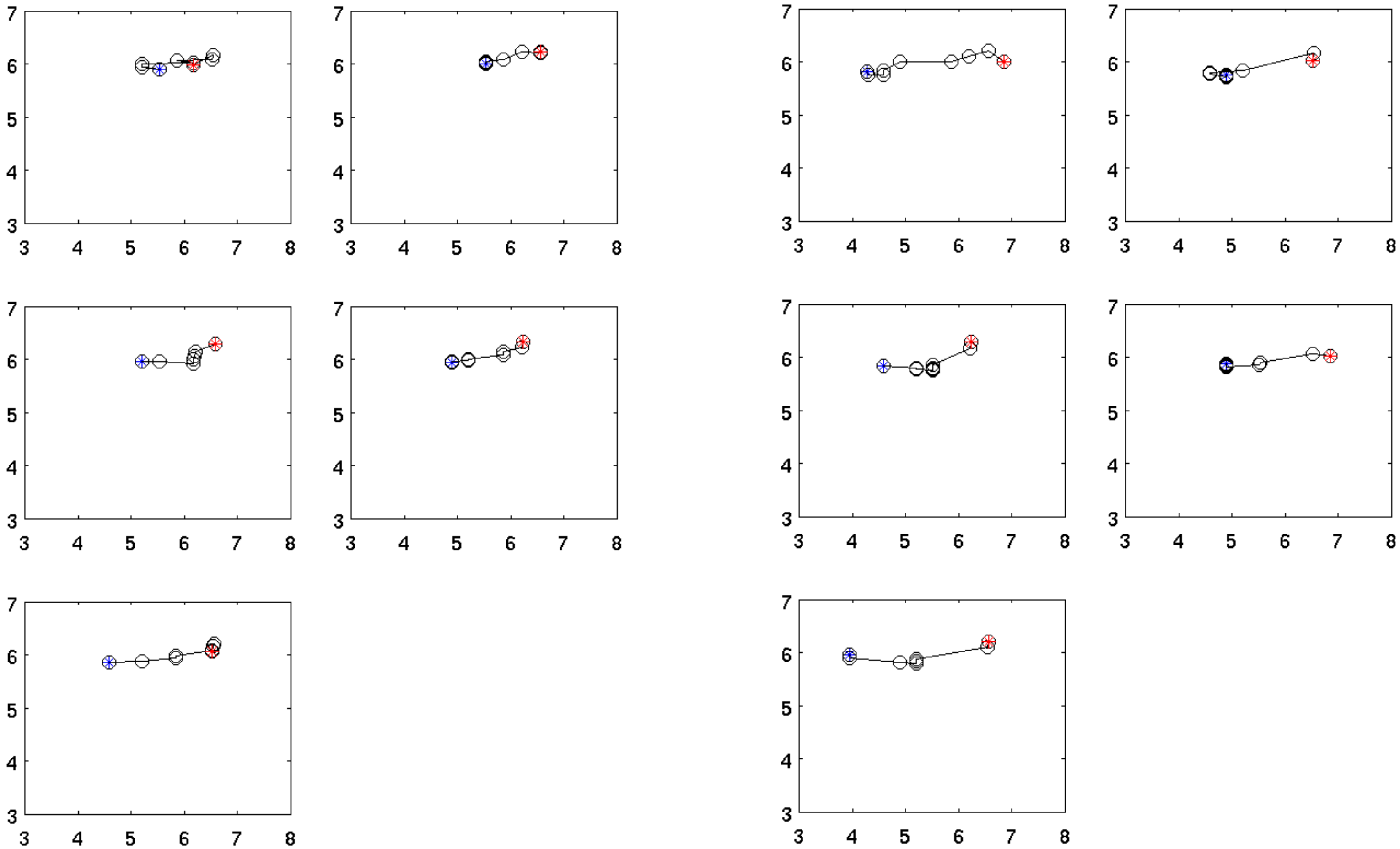


Figure 4.42: Speaker AD5's highest points on the midsagittal tongue contours over "pay" (left) and "say" (right). Scale is in cm.

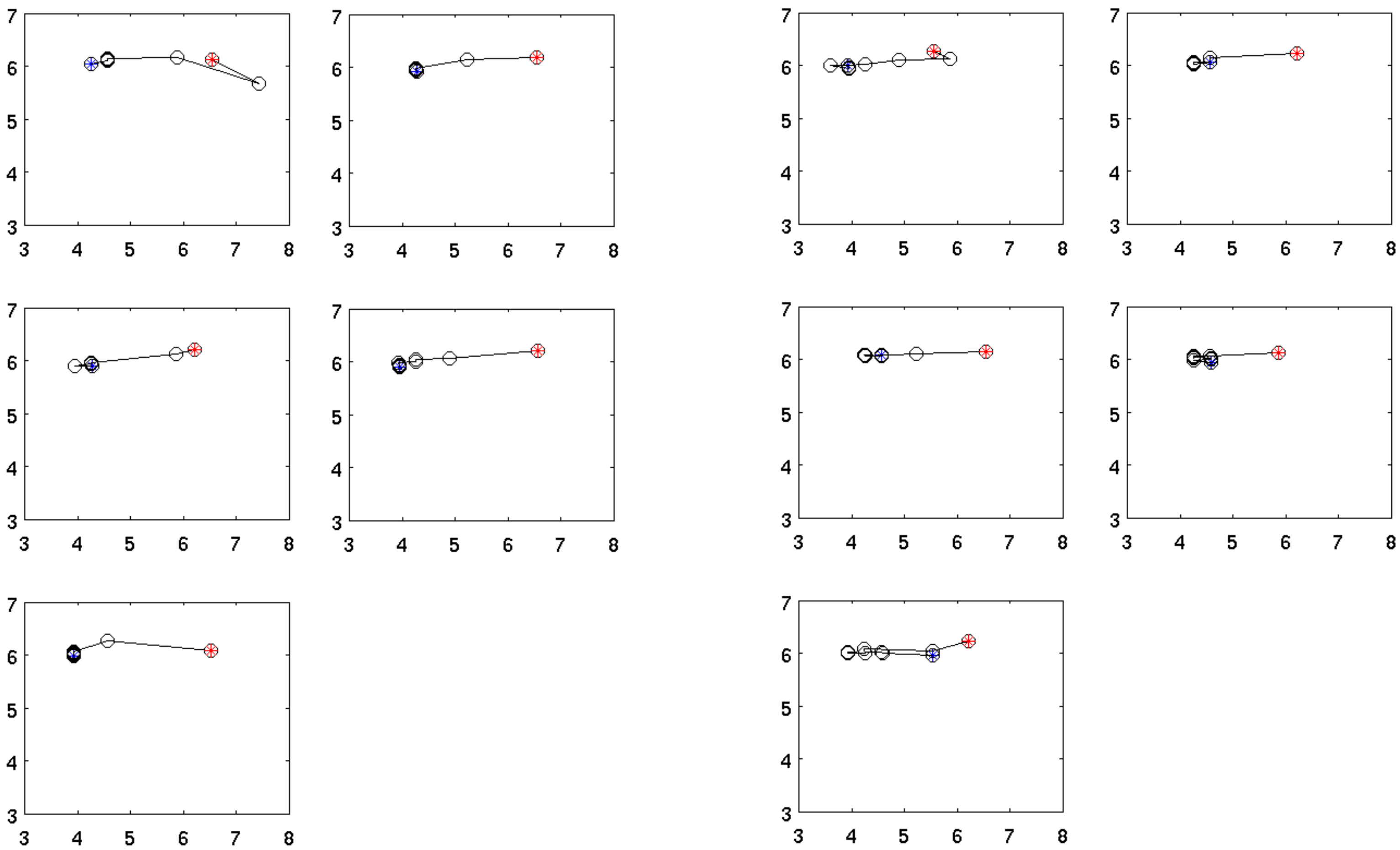


Figure 4.43: Speaker AD5’s highest points on the midsagittal tongue contours over “lay” (left) and “play” (right). Scale is in cm.

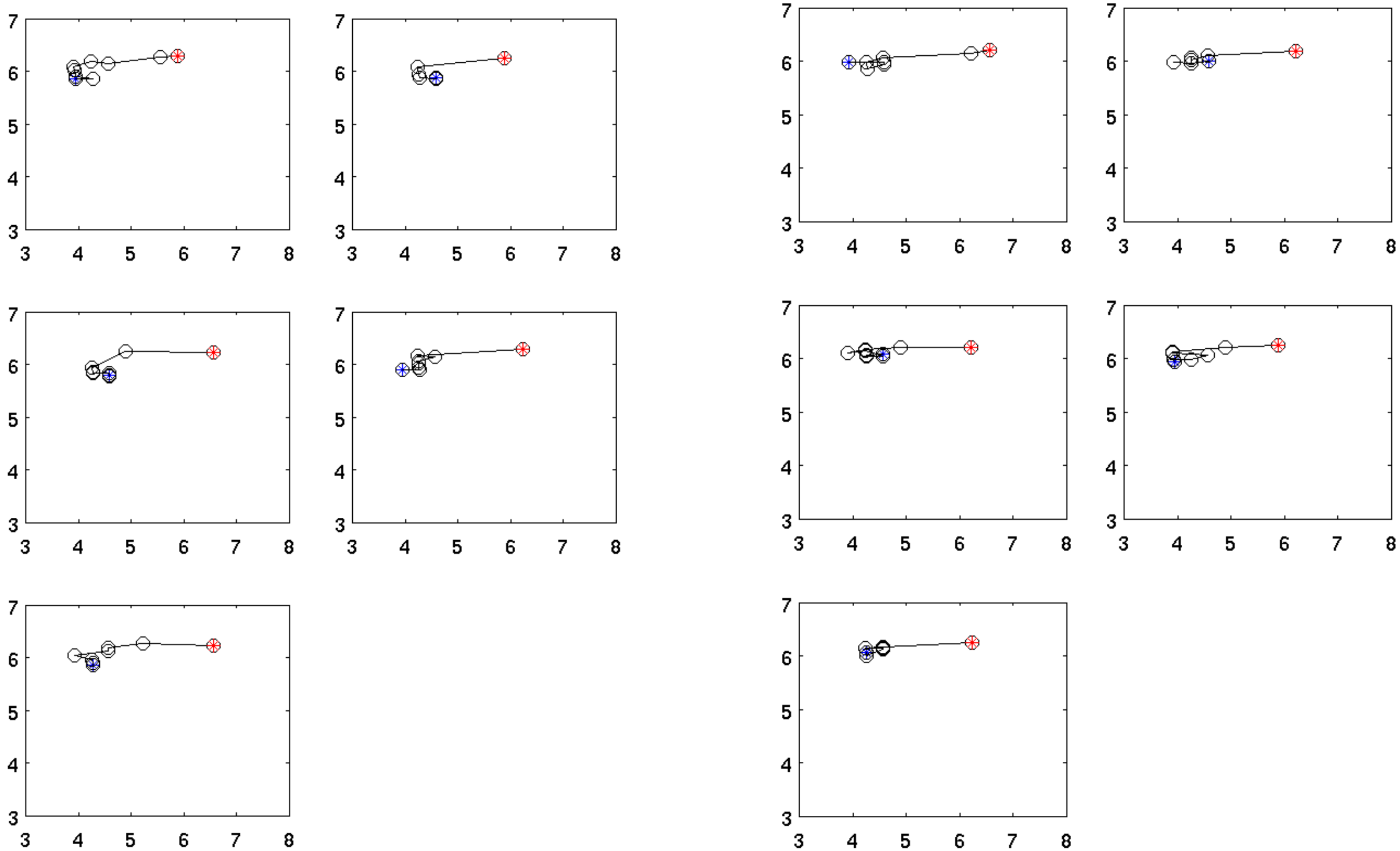


Figure 4.44: Speaker AD5's highest points on the midsagittal tongue contours over "slay" (left) and "splay" (right). Scale is in cm.

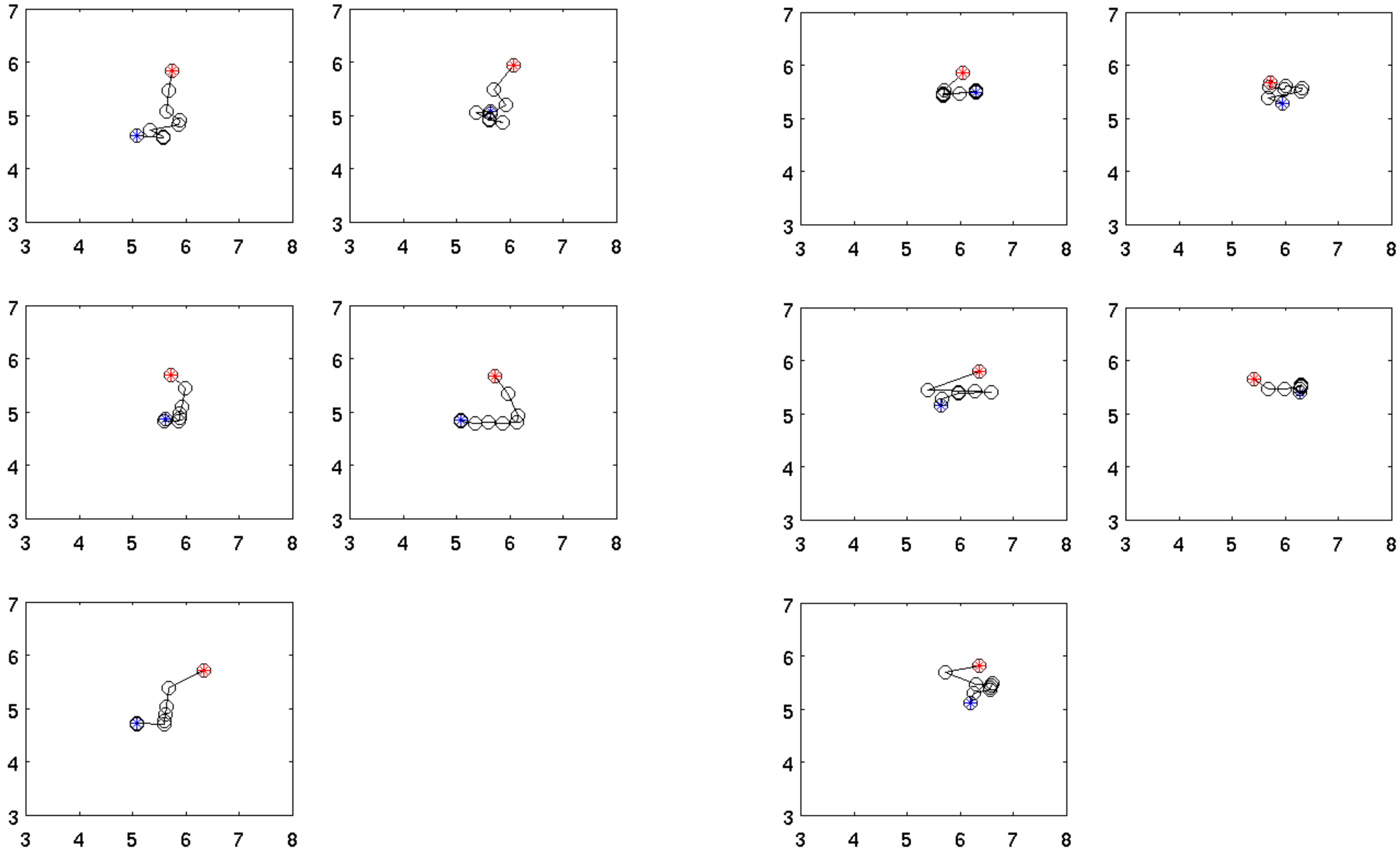


Figure 4.45: Speaker TDC1's highest points on the midsagittal tongue contours over "pay" (left) and "say" (right). Scale is in cm.

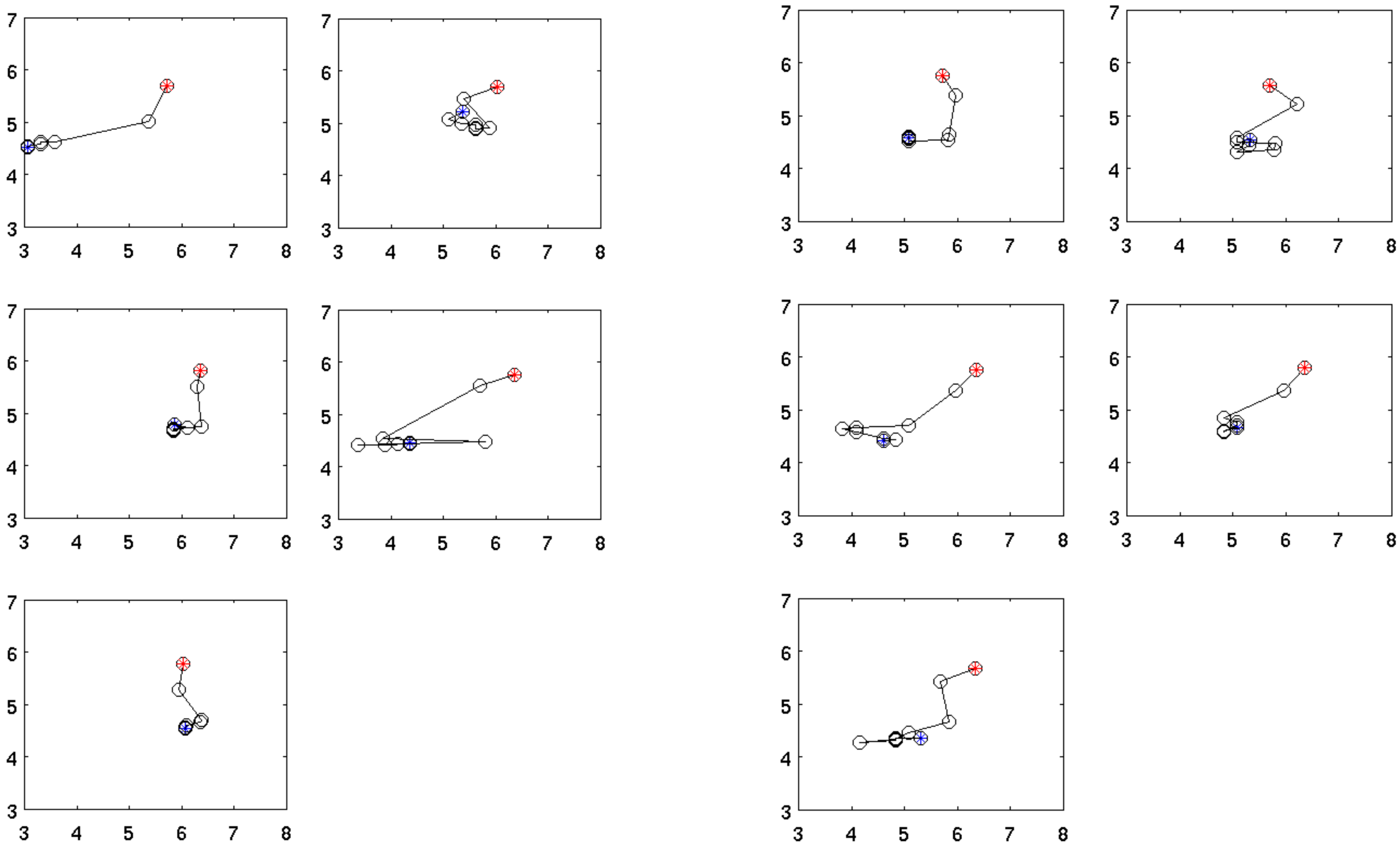


Figure 4.46: Speaker TDC1's highest points on the midsagittal tongue contours over “lay” (left) and “play” (right). Scale is in cm.

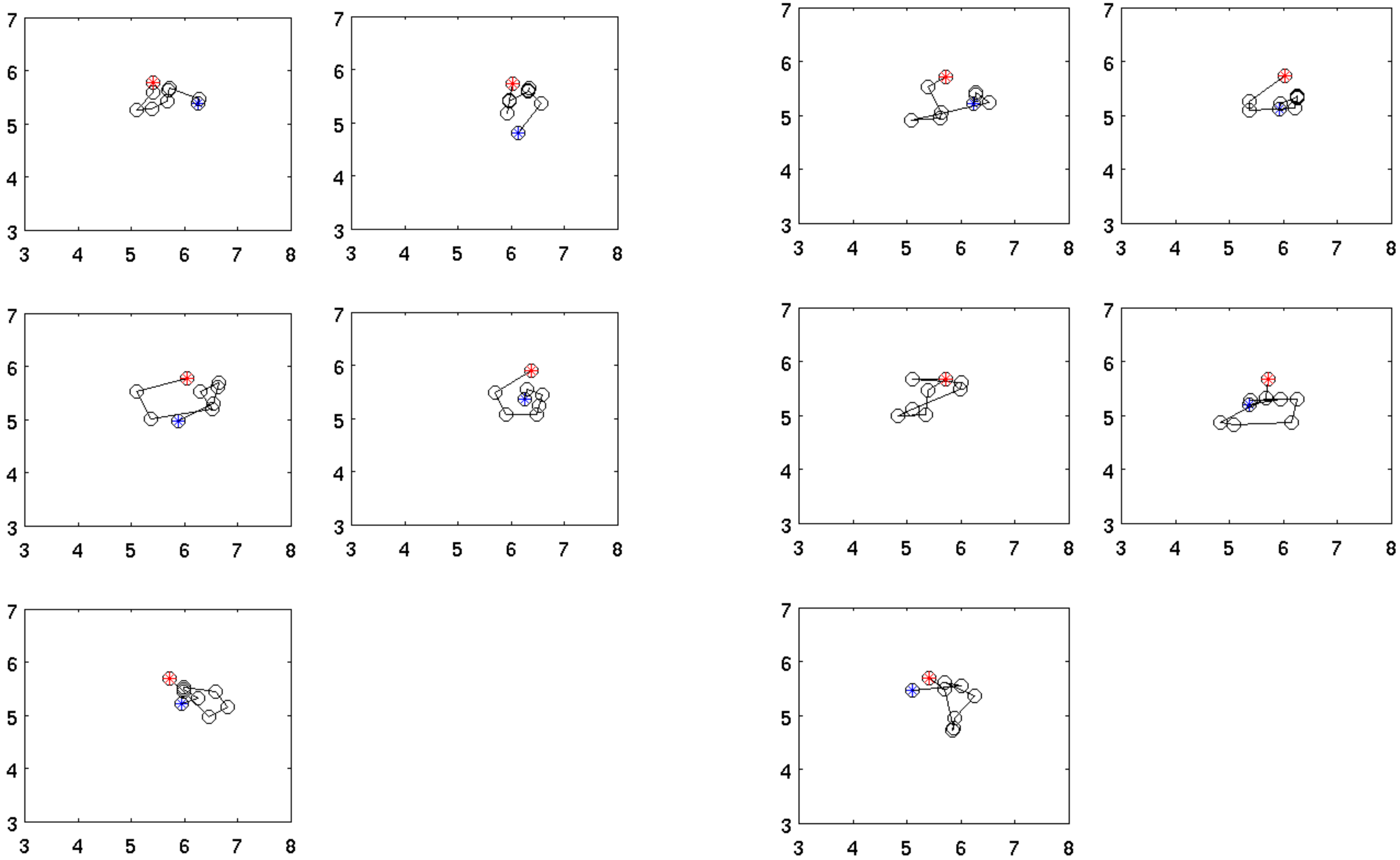


Figure 4.47: Speaker TDC1’s highest points on the midsagittal tongue contours over “slay” (left) and “splay” (right). Scale is in cm.

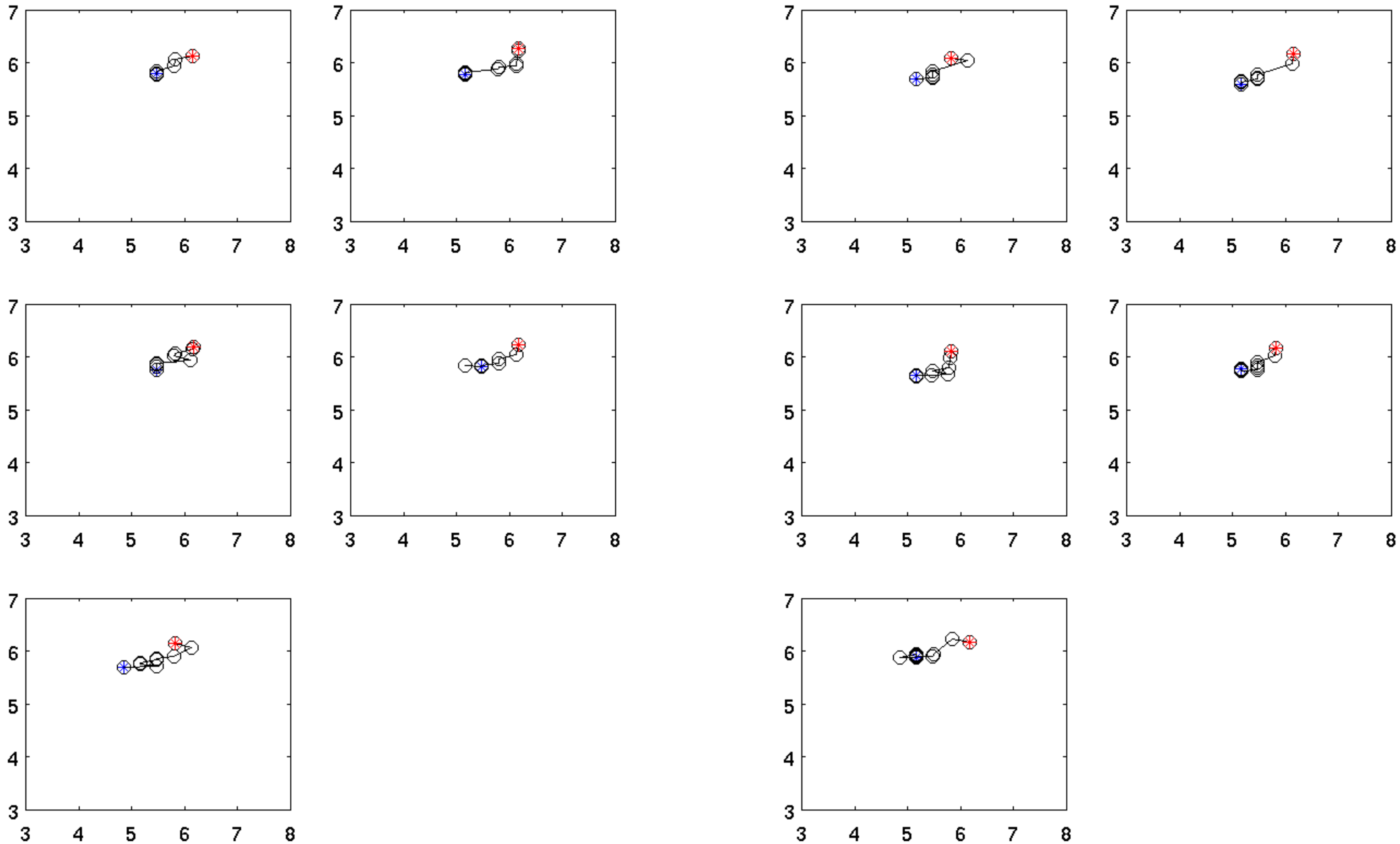


Figure 4.48: Speaker CAS1's highest points on the midsagittal tongue contours over "pay" (left) and "say" (right). Scale is in cm.

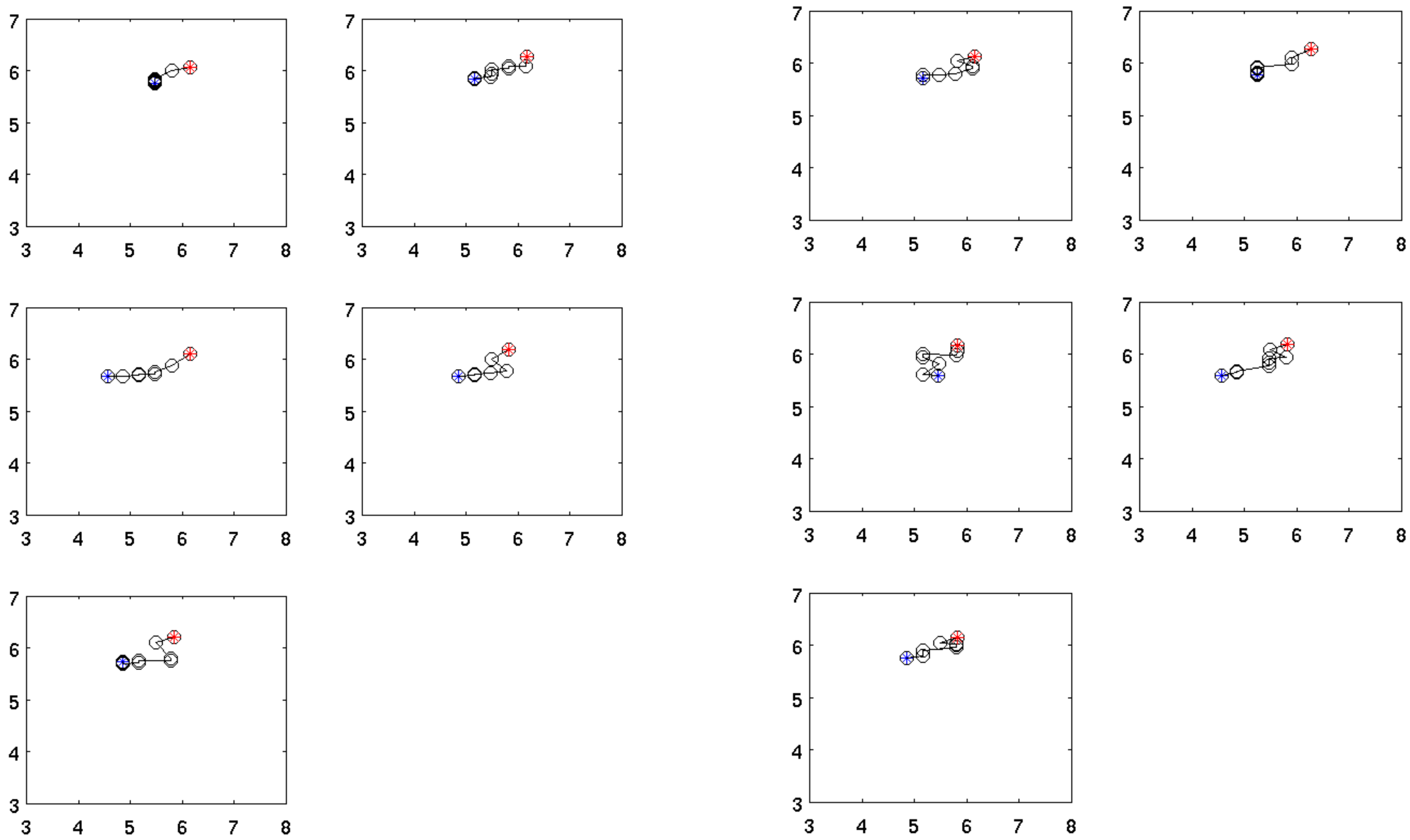


Figure 4.49: Speaker CAS1’s highest points on the midsagittal tongue contours over “lay” (left) and “play” (right). Scale is in cm.

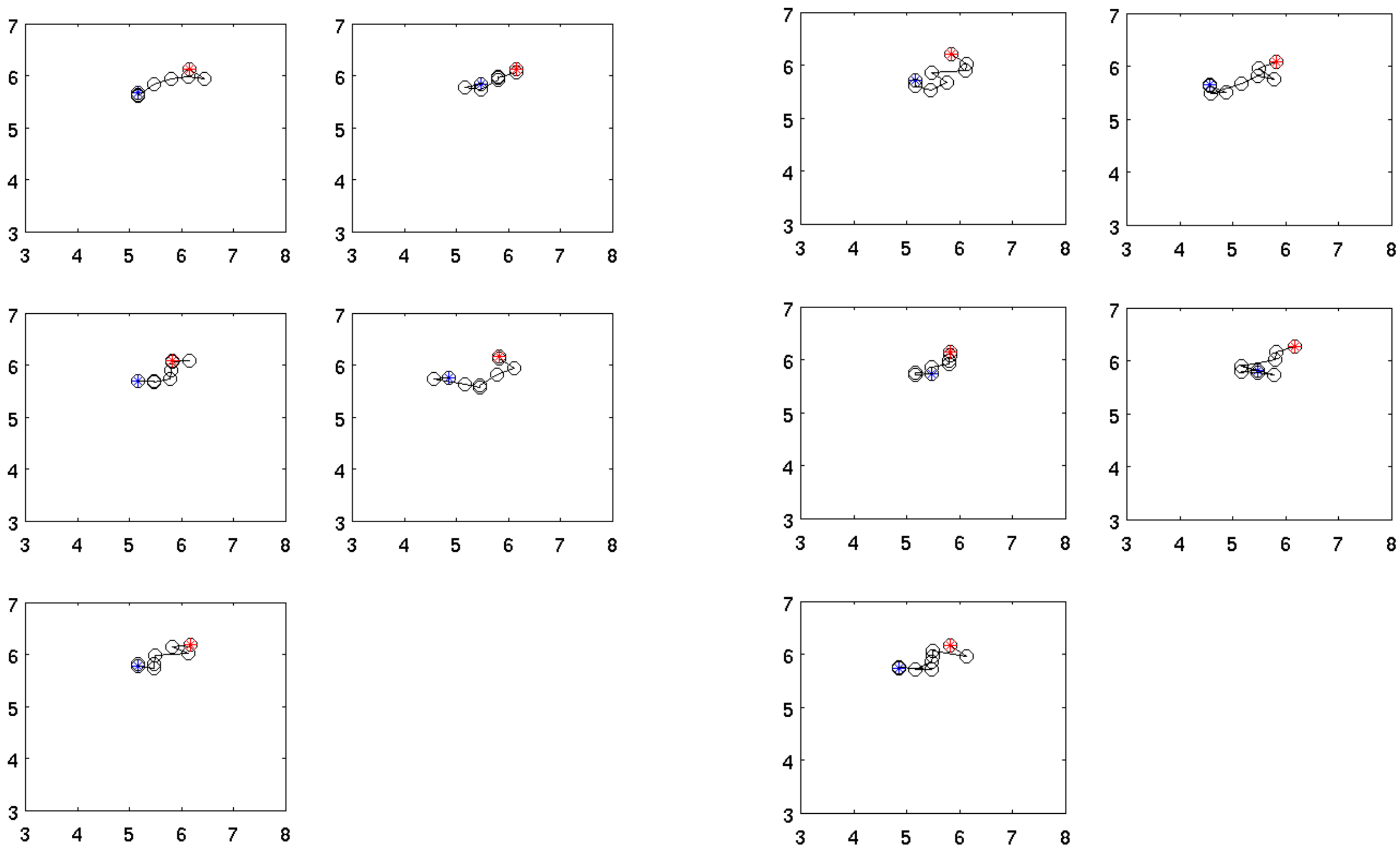


Figure 4.50: Speaker CAS1's highest points on the midsagittal tongue contours over "slay" (left) and "splay" (right). Scale is in cm.

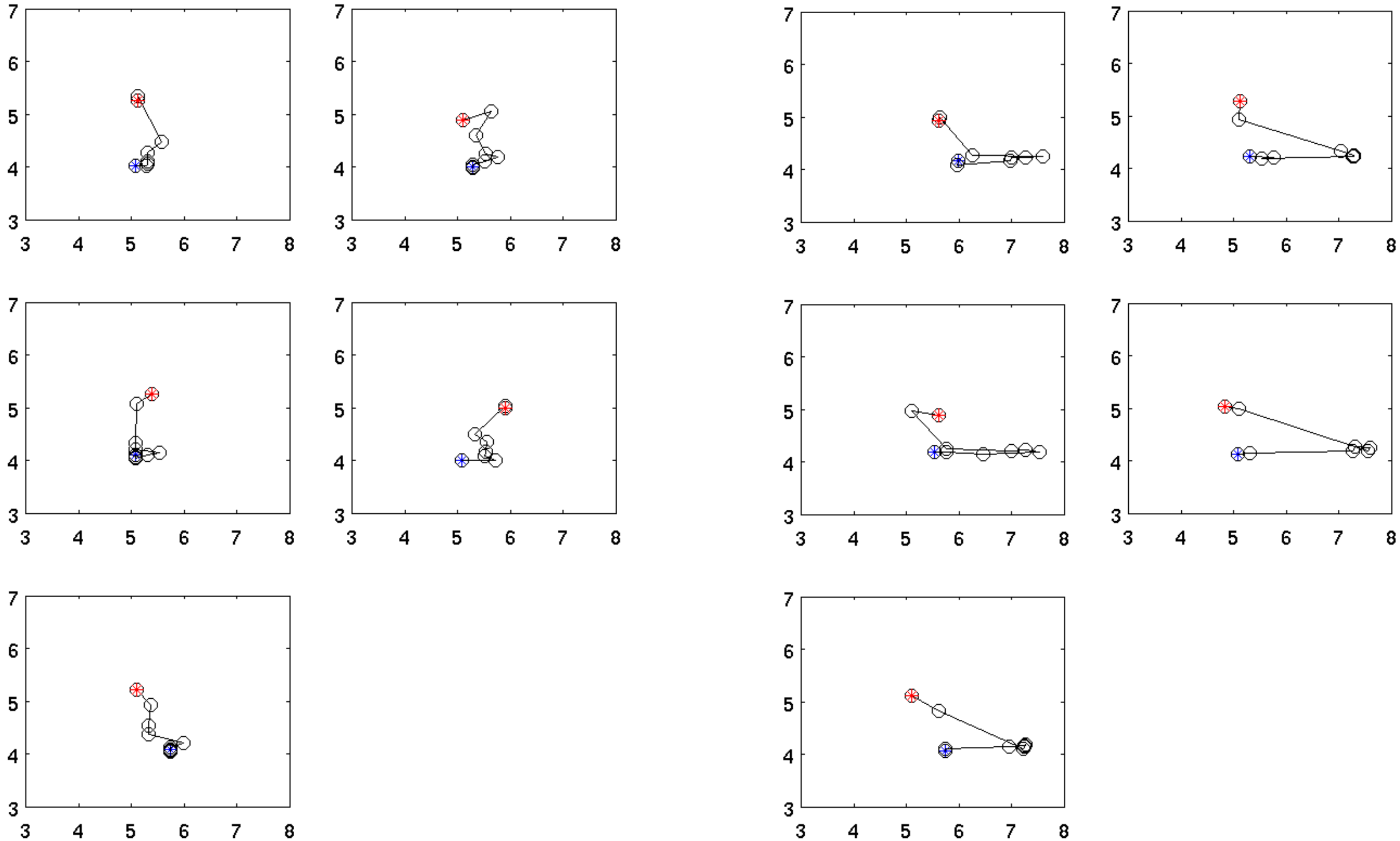


Figure 4.51: Speaker CAS2’s highest points on the midsagittal tongue contours over “pay” (left) and “say” (right). Scale is in cm.

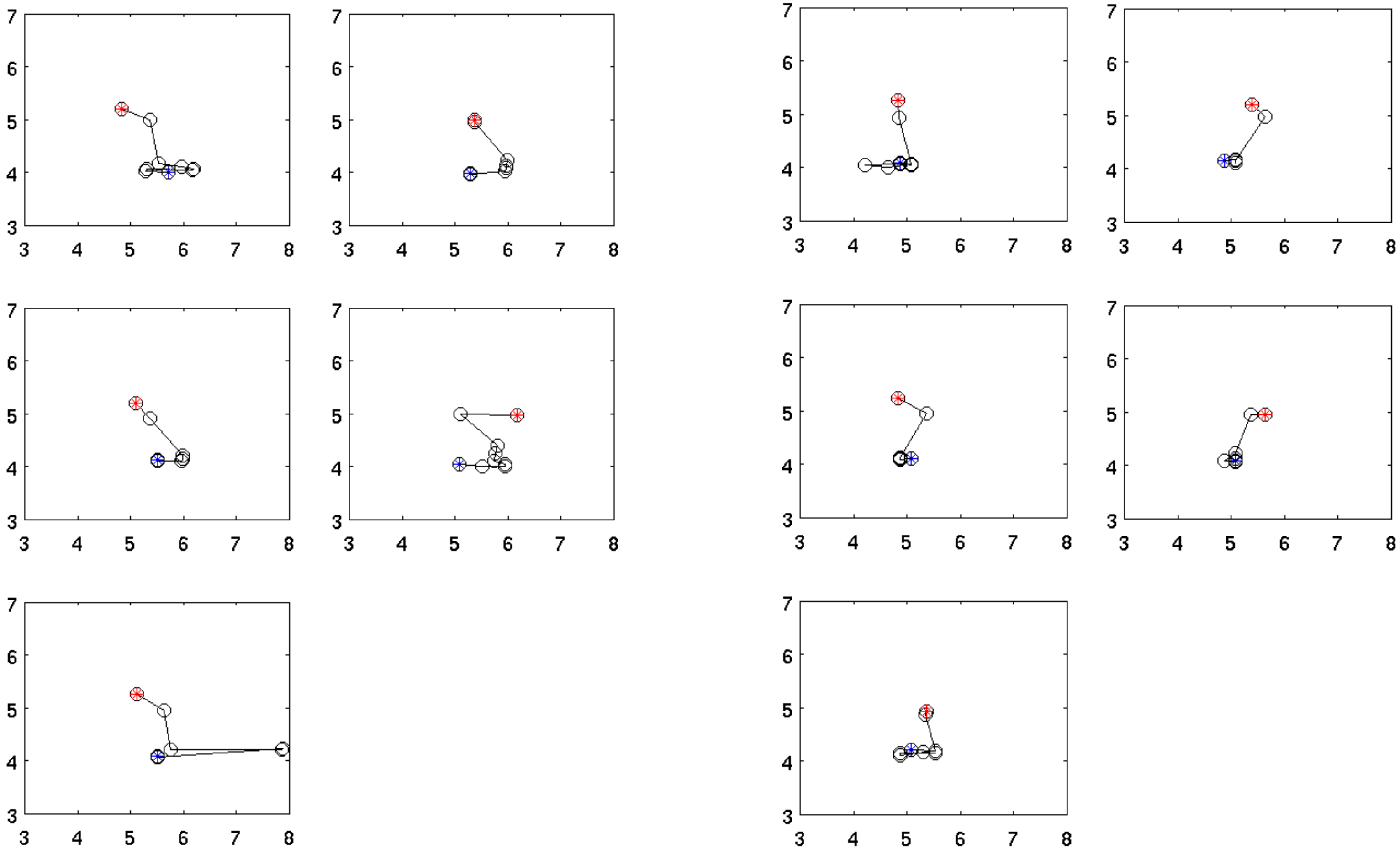


Figure 4.52: Speaker CAS2’s highest points on the midsagittal tongue contours over “lay” (left) and “play” (right). Scale is in cm.

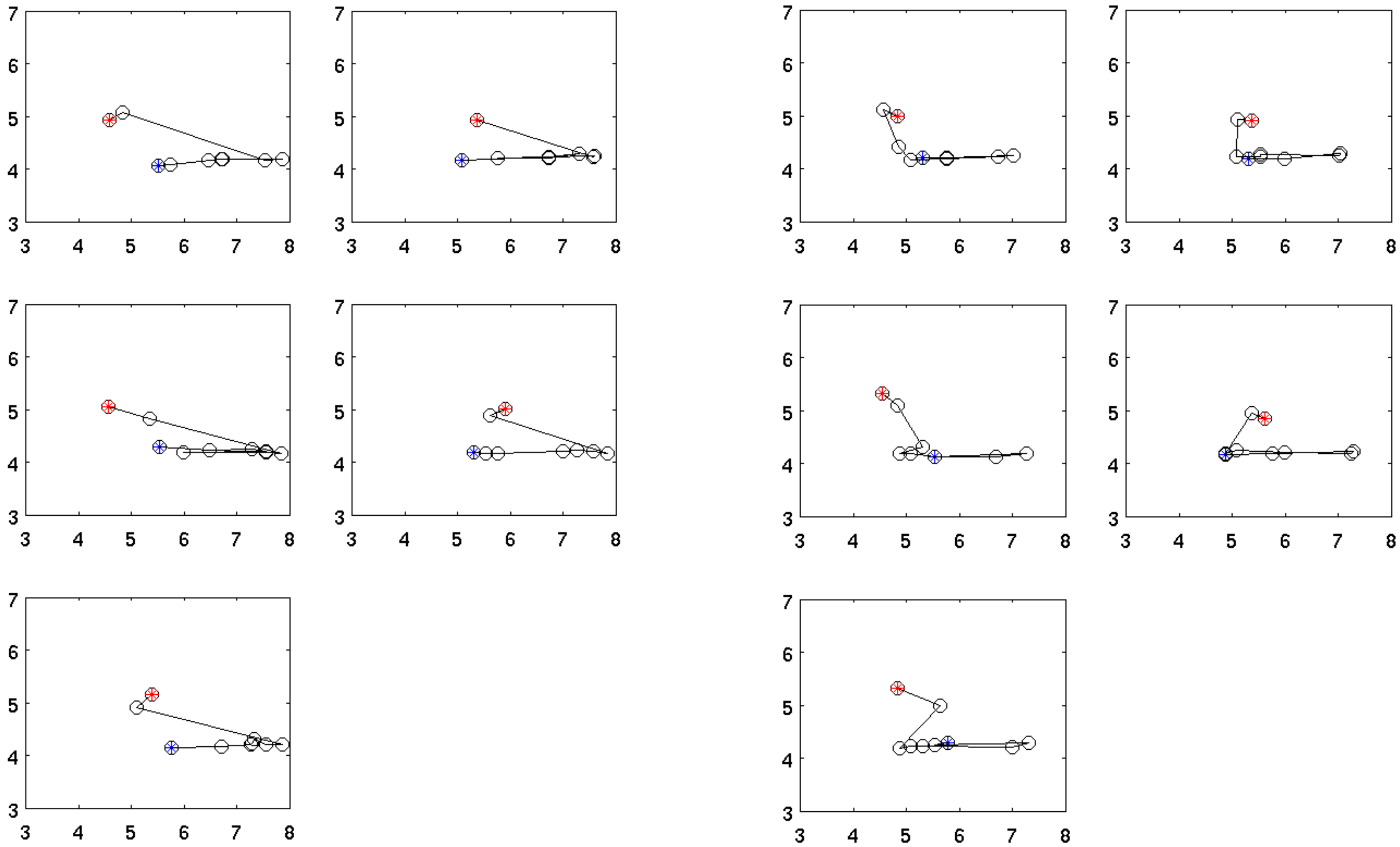


Figure 4.53: Speaker CAS2's highest points on the midsagittal tongue contours over "slay" (left) and "splay" (right). Scale is in cm.

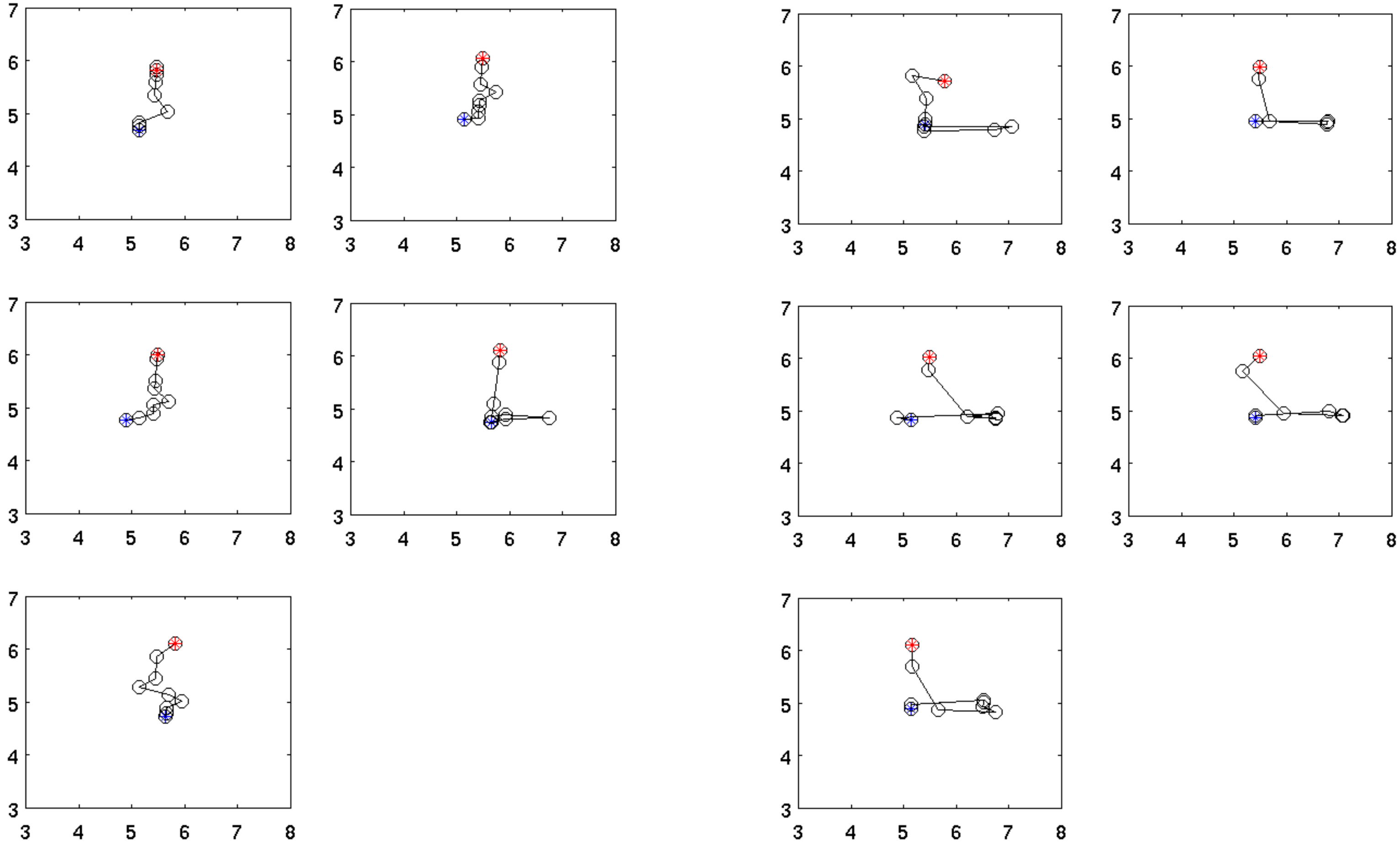


Figure 4.54: Speaker CAS3's highest points on the midsagittal tongue contours over “pay” (left) and “say” (right). Scale is in cm.

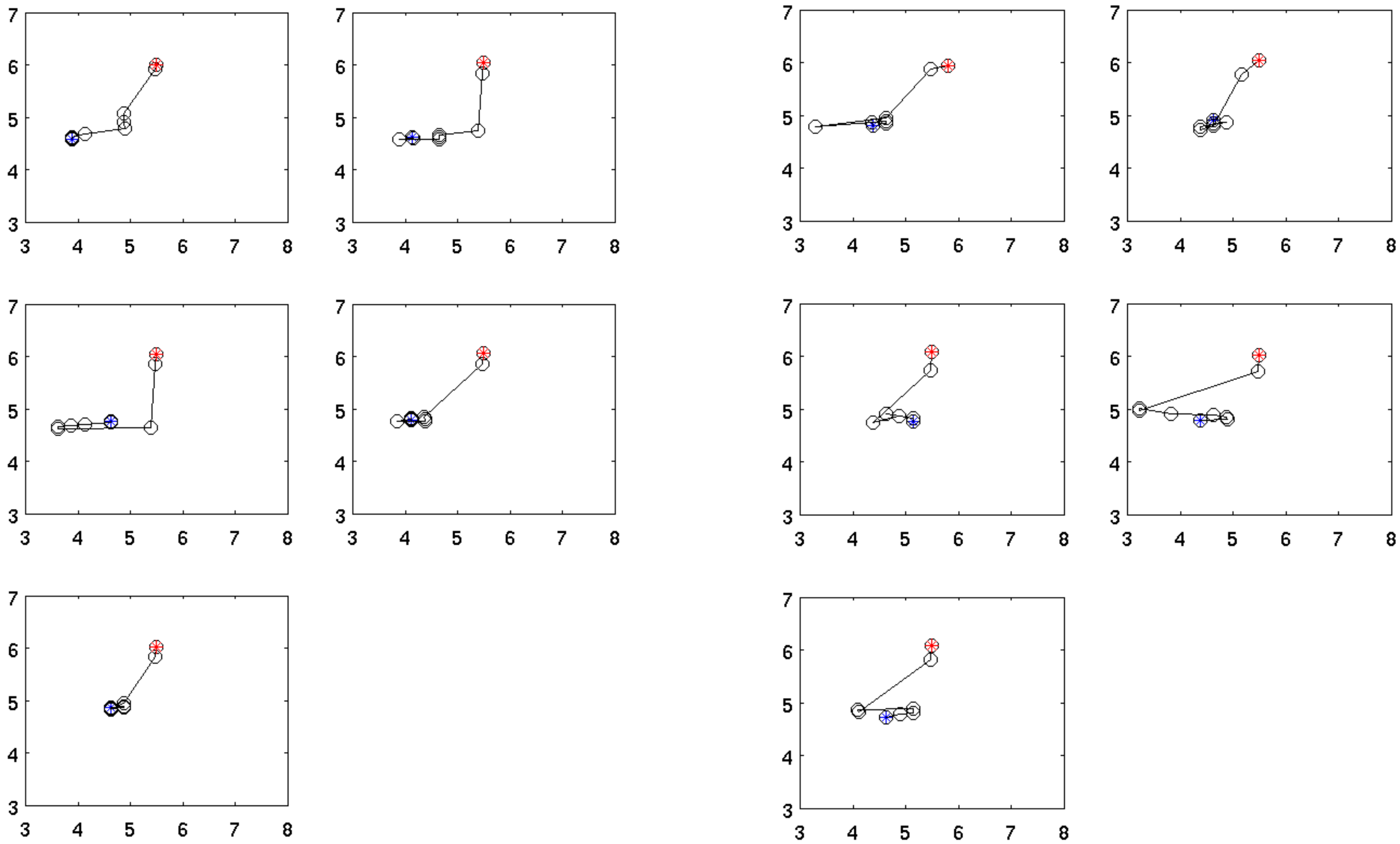


Figure 4.55: Speaker CAS3's highest points on the midsagittal tongue contours over "lay" (left) and "play" (right). Scale is in cm.

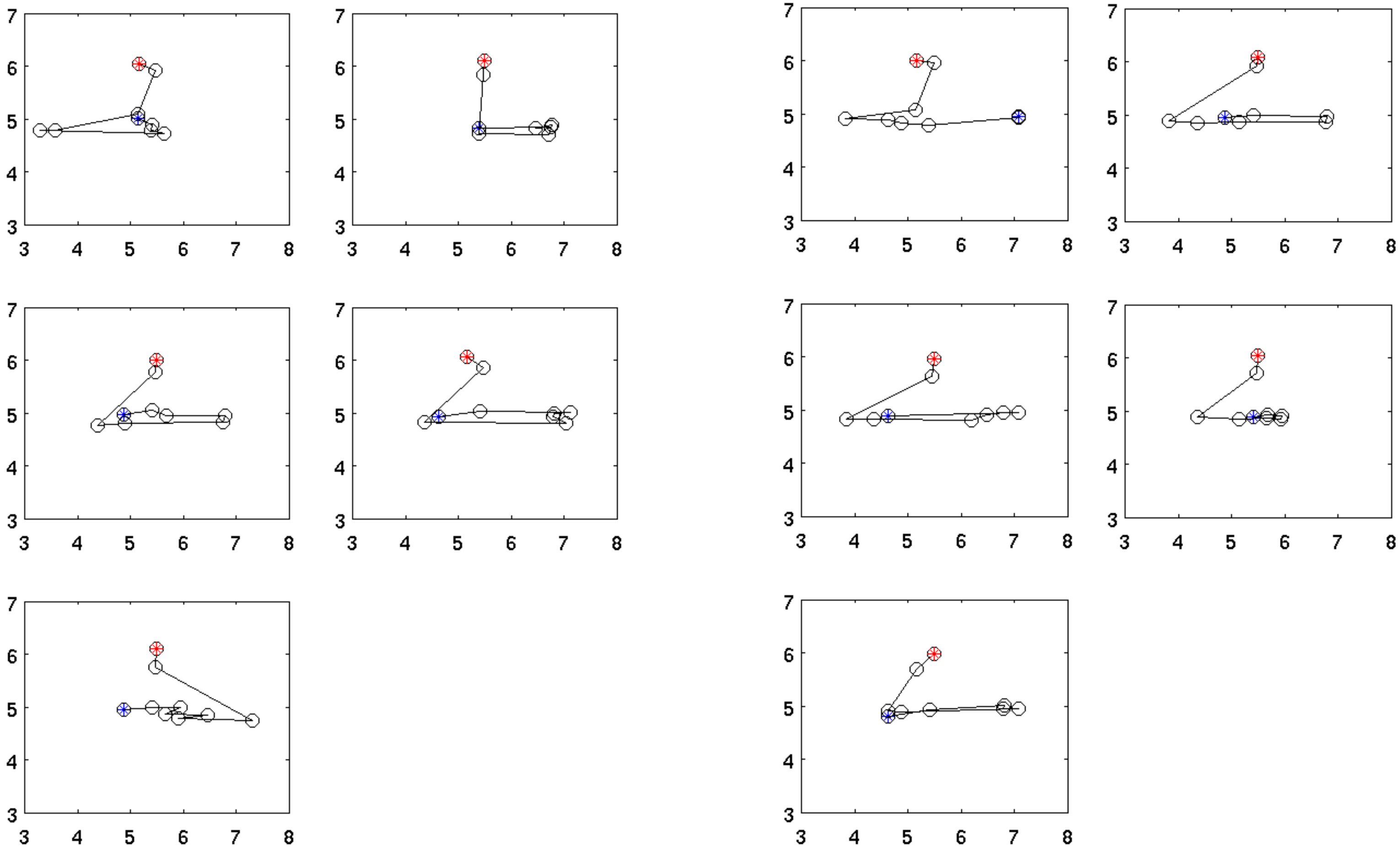


Figure 4.56: Speaker CAS3's highest points on the midsagittal tongue contours over "slay" (left) and "splay" (right). Scale is in cm.

4.5 Results of individual speakers with CAS

Because of the small number of participating speakers with CAS it is appropriate to describe them not only as a group but speaker-by-speaker as well and to compare their speech characteristics to those of the AD and the TDC speakers.

Figures 4.57, 4.58, 4.59 and 4.60 provide a visual representation of median values, IQR and range of values of syllable duration, amount of tongue movement over a syllable, rate of tongue movement over a syllable and segment duration for each of the speakers with CAS.

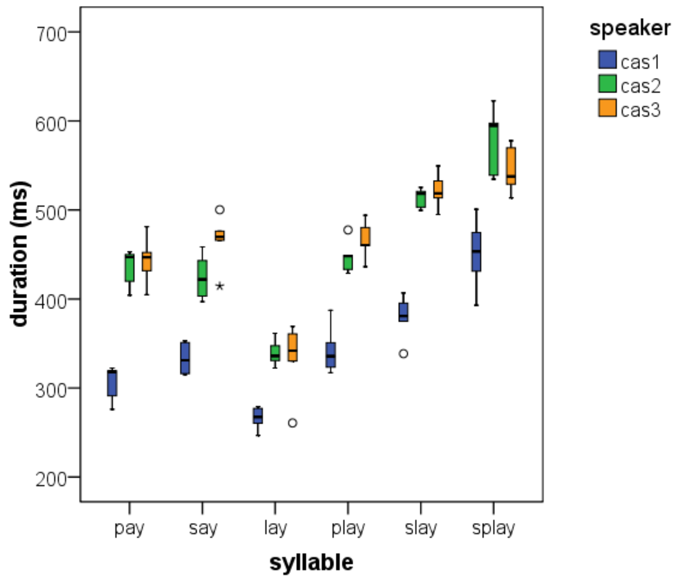


Figure 4.57: Syllable durations (ms) of each of the CAS speakers.

As can be seen, CAS1 had the shortest syllable durations, the lowest amount of tongue movement, and the slowest rate of tongue movement of the three speakers. CAS3, on the other hand, had the longest durations, the greatest amount and the fastest rate of tongue movement. Values of speaker CAS2 were in between the other two speakers, but he had similar rates of tongue movement to CAS1 and similar durations to CAS3. Similarly, CAS1 also had the shortest and CAS3 the longest segment durations. However, this was not as consistent as the other measures. CAS3 did not have the longest durations of clustered /p/ and /l/.

Overall, all three speakers increased duration with the addition of an onset segment, they all had a greater amount of tongue movement over syllables with clustered onsets than over those with singletons, and the fastest rate of tongue movement over ”lay”. Applying GLMM to the data revealed (Table 4.16) that all three speakers showed an

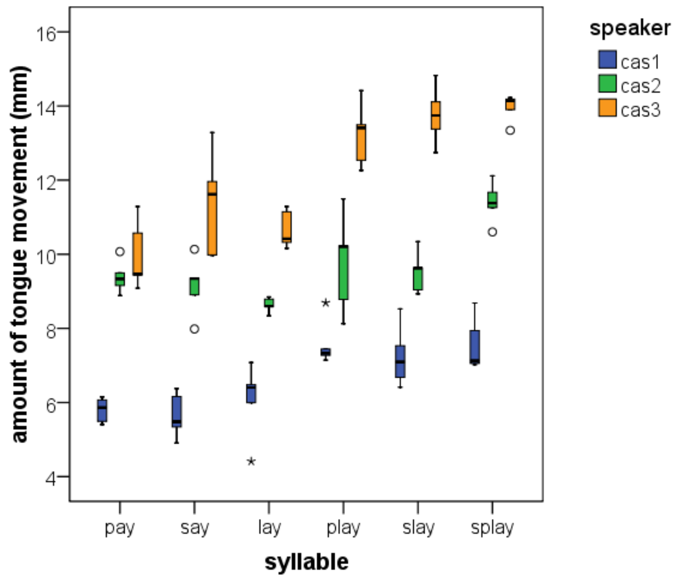


Figure 4.58: Amount of tongue movement (mm) of each of the CAS speakers.

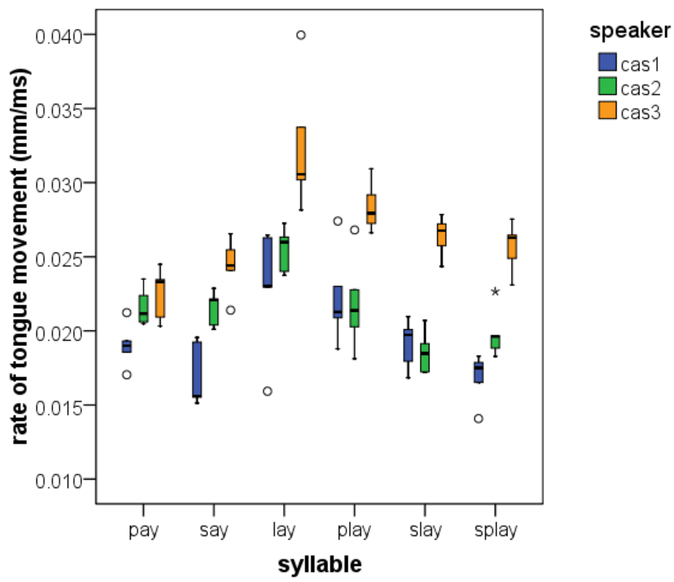


Figure 4.59: Rate of tongue movement (mm/ms) of each of the CAS speakers.

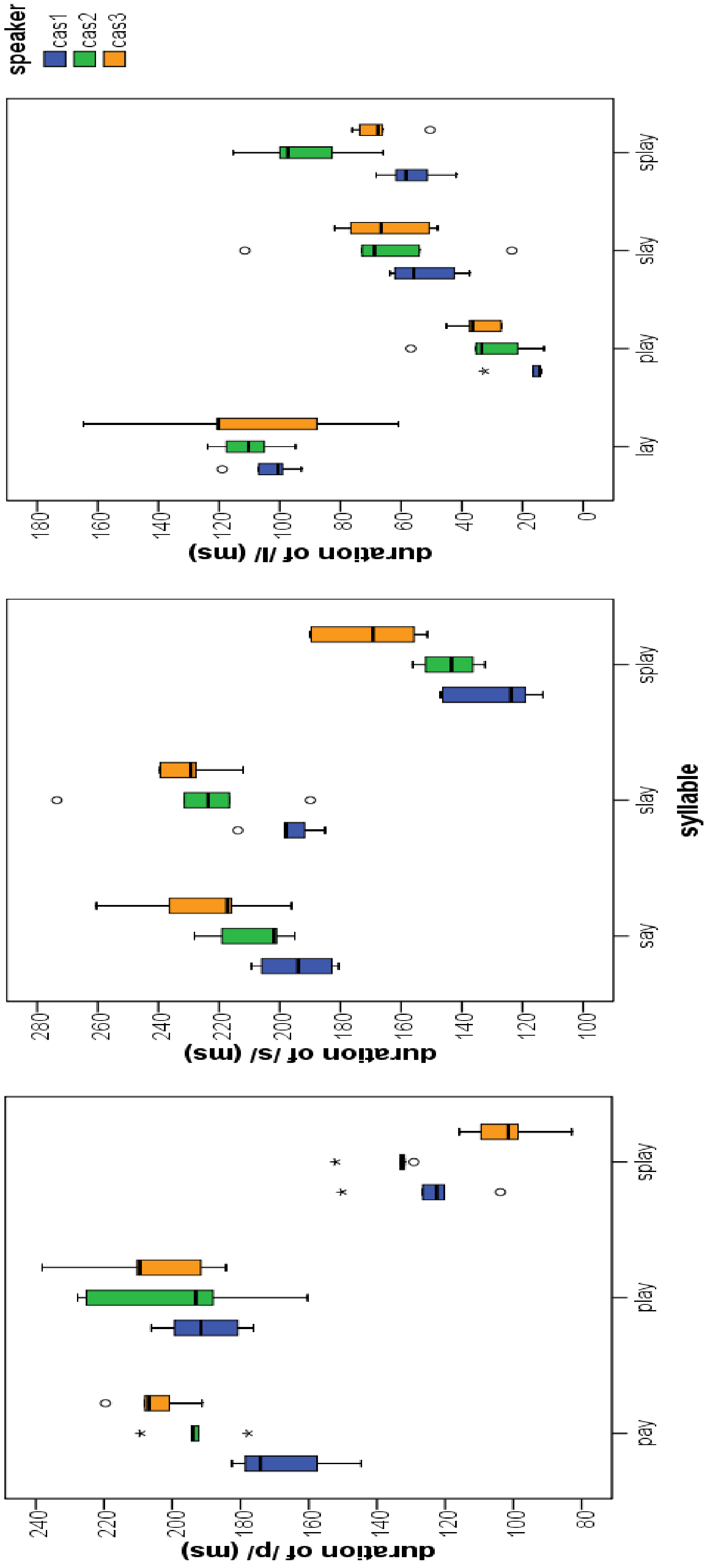


Figure 4.60: Duration (ms) of /p/ (left), /s/ (middle) and /l/ (right) as single- and clustered- onset segment of each of the CAS speakers. Note that the scale is different in each figure.

effect of the number of onset segments on duration, none of them adjusted the rate of tongue movement with respect to the number of lingual or non-lingual onset segments, and two of them, CAS1 and CAS3, also showed an effect of the number of lingual onset segments on the amount of tongue movement. The same two speakers also changed the duration of /l/ in different syllable types, but none of the three speakers adjusted duration of /p/ or /s/.

	Speaker	CAS1	CAS2	CAS3
Measure	Effect	*p*-values	*p*-values	*p*-values
Syllable duration	Number of onset segments	0.0002 *	0.0001 *	0.0026 *
Amount of tongue movement	Number of lingual onset segments	0.0348 *	0.1034	0.0080 *
Rate of tongue movement	Number of lingual onset segments	0.4230	0.1825	0.5972
	Number of non-lingual onset segments	0.7244	0.4990	0.3354
/p/ duration	Type of syllable onset	0.6462	0.3036	0.1380
/s/ duration		0.2154	0.4232	0.3414
/l/ duration		0.0052 *	0.1038	0.0342 *

Table 4.16: *p*-values obtained by GLMM modelling of the effect of the number of onset segments on syllable duration, of the number of lingual onset segments on the amount of tongue movement, of the number of lingual and non-lingual onset segments on the rate of tongue movement and of the type of syllable onset (single segment or cluster) on segment duration for each of the CAS speakers. *p*-values < 0.05 are marked by (*) and indicate significance.

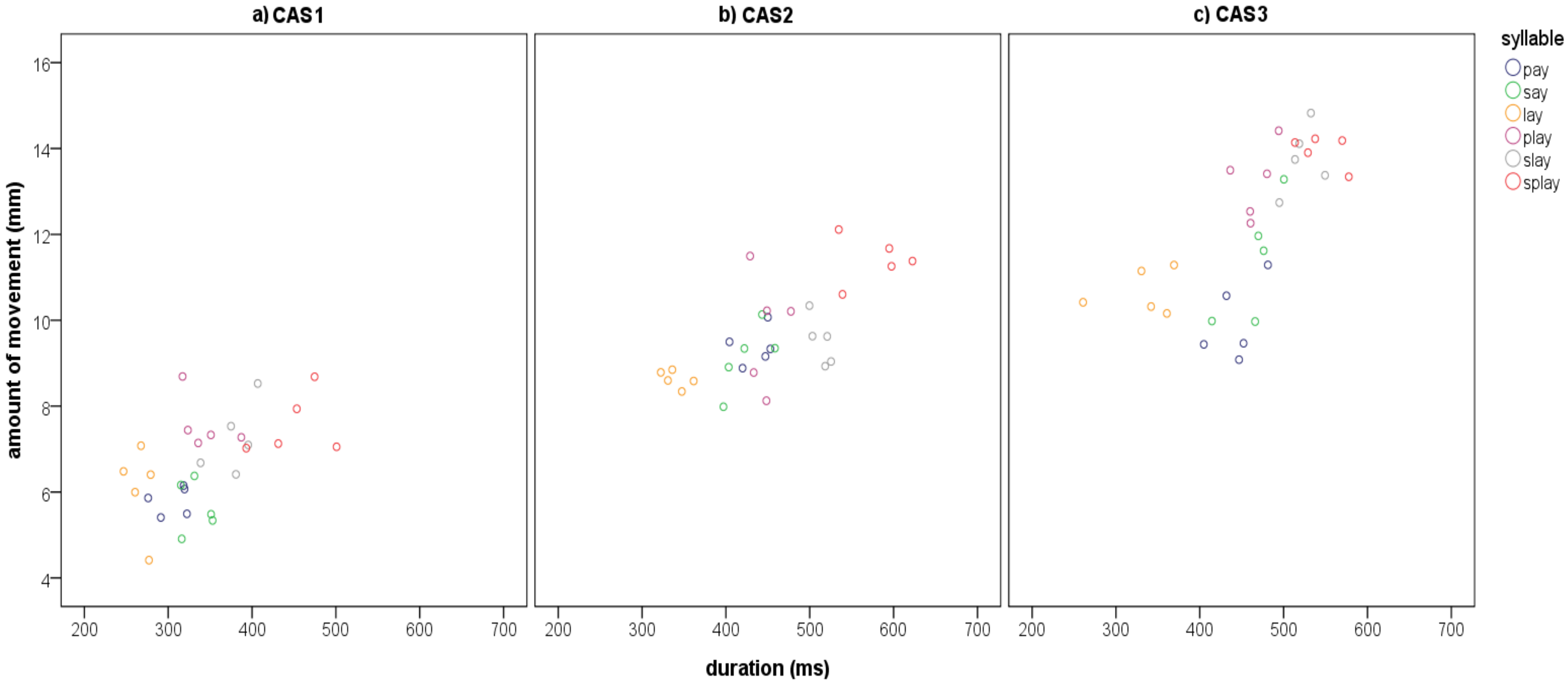


Figure 4.61: Distribution of duration (ms) and amount of tongue movement (mm) over the six syllables for each of the CAS speakers: (a) CAS1, (b) CAS2, (c) CAS3

Differences between speakers in the relation between syllable duration and the amount of tongue movement can also be observed in Figure 4.61. It shows again that CAS1 had shorter durations and a smaller amount of tongue movement over all syllable types than the other two speakers. But this figure also shows that distribution of syllables is very similar across the speakers. "lay" seems to be separated from other targets, "pay" and "say" occupy the same area in the distribution, and syllables with clustered onsets are more similar to each other, with "splay" having the highest measures of duration and amount of tongue movement.

When comparing measures of individual speakers with CAS to those of the control speakers they again showed some differences in syllable duration, amount and rate of tongue movement. Figures 4.62 - 4.67 show the ranges between minimum and maximum values (represented by bars) and median values (represented by circles) for the three speakers with CAS and for either ten AD speakers (Figures 4.62, 4.64 and 4.66) or ten TDC speakers (Figures 4.63, 4.65 and 4.67) on each of these three measures.

In the case of syllable duration, speaker CAS1 produced durations that are comparable to those of adults and fit in the middle of the ranges produced by individual AD speakers, while values of CAS2 and CAS3 are higher and similar only to the two AD speakers with the longest durations (Figure 4.62). Similar division between CAS1 and the other two speakers can be seen when comparing their syllable durations to those of the TDC speakers (Figure 4.63). While durations of CAS1 fall into the lower part of the TDC speakers' ranges, durations of CAS2 and CAS3 occupy the same range as most of the TDC speakers.

Even more differences between the CAS speakers can be observed in amount of tongue movement. When compared to both the AD speakers (Figure 4.64) and the TDC speakers (Figure 4.65), CAS1 produced less movement over the syllables with lower minimum and median values than control speakers on almost all syllables, minimum -maximum ranges of speaker CAS2 were comparable to most of control speakers and CAS3 was again similar only to control speakers with the highest values.

Speaker CAS1 additionally displayed slower rates of tongue movements than most of the AD speakers, while all values of CAS2 and CAS3 were inside the values of individual AD speakers (Figure 4.66). CAS2 and CAS3, however, differed in that values of the former occupied the lower range of the AD measures while the values of the latter fitted in the middle part. When compared to the TDC speakers, none of the speakers with CAS produced slower or faster rates than the control speakers, but they did show some differences from each other. CAS1 had the slower rates on almost all syllables and CAS3 the fastest. Overall, CAS1 and CAS2 produced more similar rates of tongue movement than CAS3.

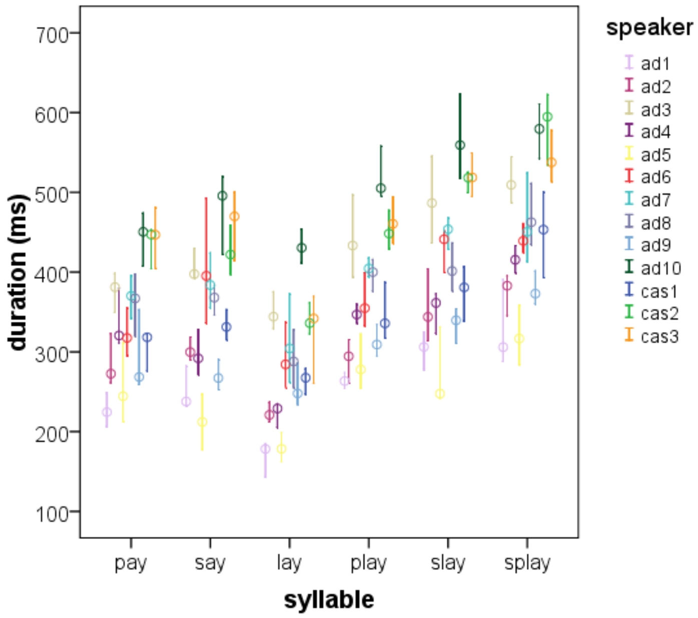


Figure 4.62: Minimum - maximum range (bars) and median values (circles) of syllable duration (ms) of the three CAS and ten AD speakers.

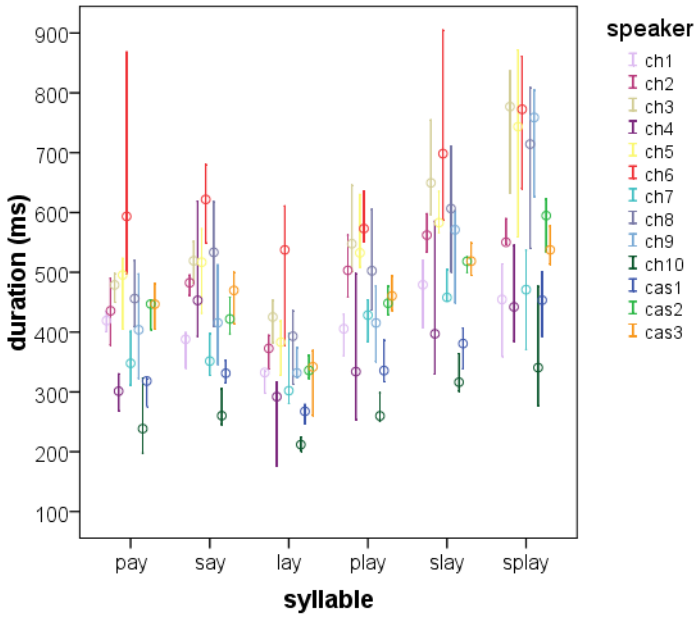


Figure 4.63: Minimum - maximum range (bars) and median values (circles) of syllable duration (ms) of the three CAS and ten TDC speakers.

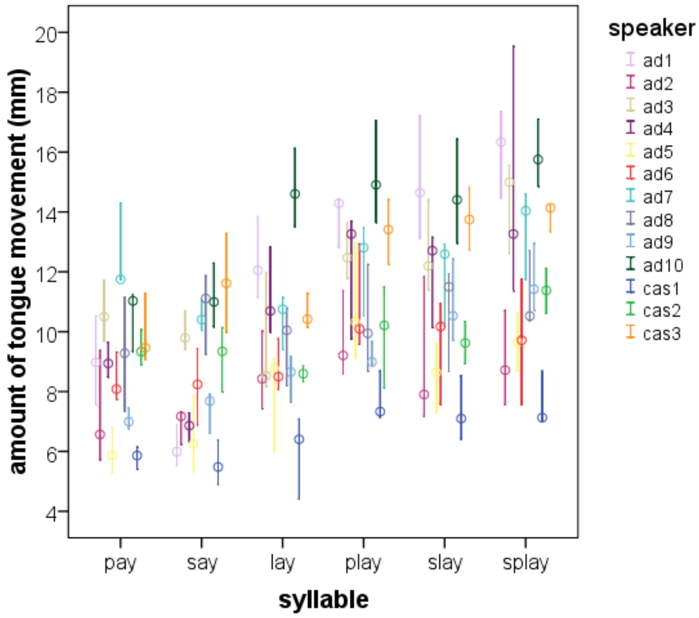


Figure 4.64: Minimum - maximum range (bars) and median values (circles) of amount of tongue movement (mm) of the three CAS and ten AD speakers.

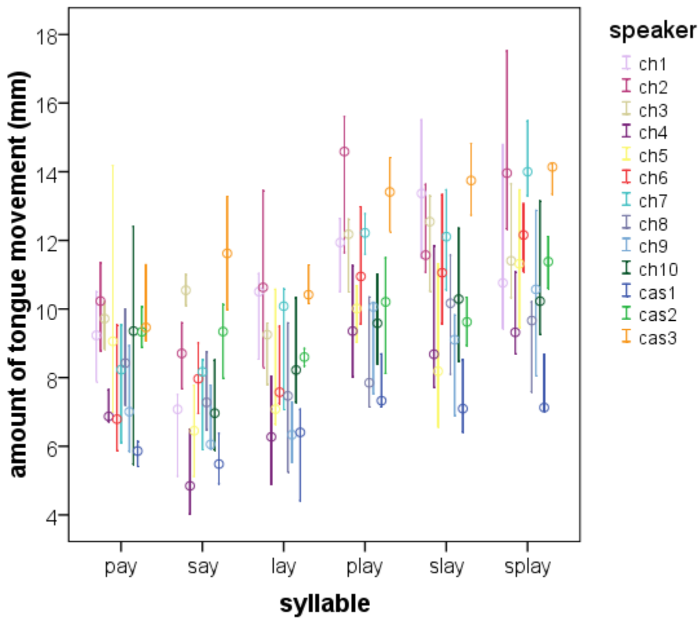


Figure 4.65: Minimum - maximum range (bars) and median values (circles) of amount of tongue movement (mm) of the three CAS and ten TDC speakers.

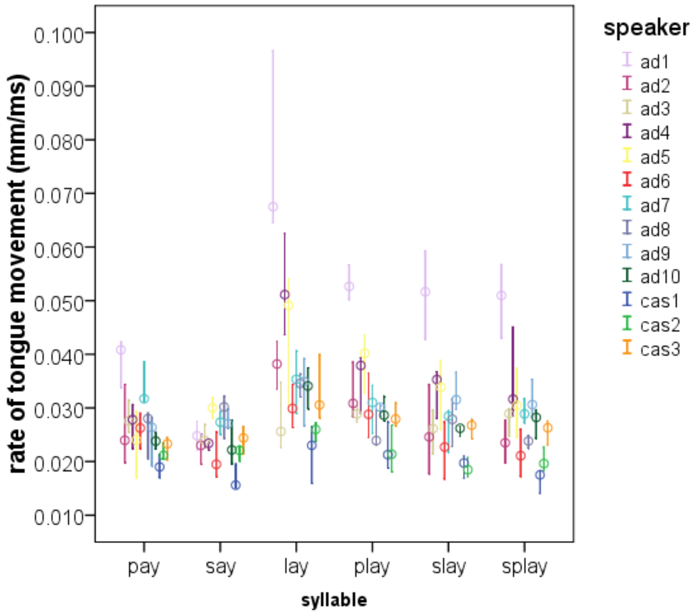


Figure 4.66: Minimum - maximum range (bars) and median values (circles) of rate of tongue movement (mm/ms) of the three CAS and ten AD speakers.

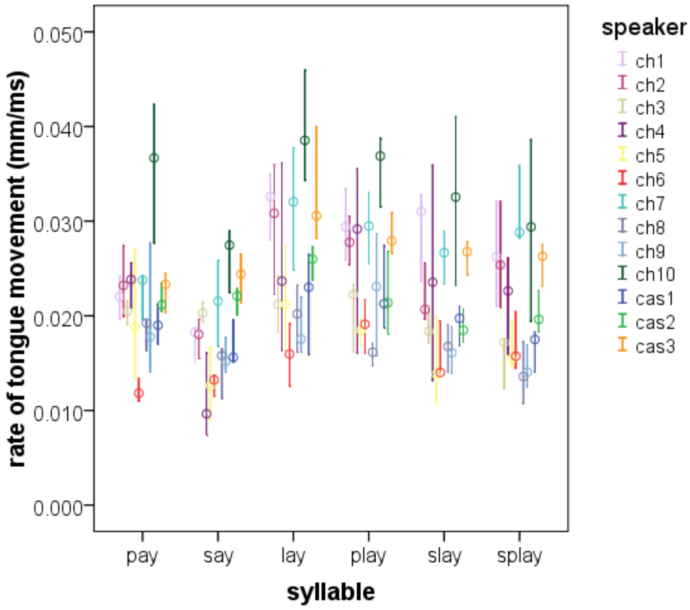


Figure 4.67: Minimum - maximum range (bars) and median values (circles) of rate of tongue movement (mm/ms) of each of the three CAS and ten TDC speakers.

Chapter 5

Discussion

The research presented in this thesis aimed at revealing speech characteristics of CAS by analysing acoustic and articulatory data. Speech of speakers with CAS was compared to mature productions of adults and immature productions of typically developing children. The results suggest first, that speakers with CAS can be differentiated from the two control groups on some of the measures but not on all of them and second, that ultrasound is a valuable methodology for investigating CAS because it allows direct observation of tongue movements during speech and because it enables tracking changes in articulatory and acoustic domains at the same time.

The discussion is organized in five main sections. First, statistical procedures applied to the data are discussed. The second part discusses results of the AD and the TDC groups, and the third of the CAS group compared to the control groups and of individual speakers with CAS compared to individual AD and TDC speakers. Finally, comments are made about some methodological issues of the study and possible future research.

5.1 Comparing statistics

One of the challenges of this thesis turned out to be choosing appropriate statistical methods. All data was first analysed with the standard non-parametric procedures, chosen because of the requirements of the data (small number of measures, non-normal distribution) and their application in similar previous studies. However, because of the multiple pair tests performed to answer research questions, these methods raised another problem, namely that of adjusting the alpha level. As discussed in Chapter 3.5, the amount of adjustments, or even the necessity to adjust the alpha level, is still not agreed (Saville, 1990; Cabin and Mitchell, 2000; Nakagawa, 2004). For this reason, the decision was made to apply the most frequently used Bonferroni correction and to use the least conservative approach of adjusting p-values only for the number of tests performed on a data set (e.g., testing for the effect of number of syllable onset segments on syllable duration inside a speaker group was adjusted only for the six possible syllable pairs, resulting in the alpha value reducing from 0.05 to 0.008). Applying a more strict adjustment, e.g., adjusting for all the 183 tests performed in the study, would reduce the acceptable alpha level to 0.00027, and make it very unlikely that any of the tested pairs would show a significant difference between them. However, when observing the recordings and inspecting the data, some differences between speaker groups were apparent and it seemed likely to be wrong to discard them as not significant just because of the strict adjustment to the alpha level. For example, although using the least strict approach to Bonferroni corrections revealed some differences between the groups, the doubt remained about the correctness of the level of adjustments and about the suitability of adjustments for this particular set of data.

For this reason, different methods were explored and eventually it was decided to apply GLMM as well. GLMM have been only recently applied to linguistic data but as explained in Baayen (2008) and Baayen et al. (2008) and summarised in Chapter 3.5 it is a more appropriate method than the standard ANOVA (and corresponding non-parametric methods) because it models speakers and test materials as random variables and is in that way less tied to the specific speakers or speech material used, it can cope with unbalanced data sets and data that is not normally distributed. The method seemed an appropriate statistical procedure for the data and the effects of syllable structure and speaker groups on the measures were additionally tested by applying GLMM.

Table 5.1 presents the outcome of the hypothesis testing by applying non-parametric and GLMM methods. In the case of 39 tested assumptions, non-parametric methods confirmed the hypotheses in 14 cases and GLMM in 17. Neither of the methods confirmed even half of the predictions. It is tempting to conclude that both methods have equal power of revealing the differences between the speaker groups. But this would not be a valid conclusion. It is important to stress that in 14 cases the two methods lead to opposite conclusions: effect of the number of onset segments on syllable duration in the CAS group (see Hypothesis 1 in section 2.6), changes in the duration of /l/ in all three speaker groups (see Hypothesis 2 in section section 2.6), differences in the syllable duration between the AD and the TDC groups, and between the AD and the CAS groups (Hypothesis 3), difference in the variability of the syllable duration between the

AD and the TDC groups (see Hypothesis 4 in section 2.6), effect of the number of lingual segments on the amount of tongue movement in the AD and the TDC groups (see Hypothesis 5 in section 2.6), effect of the number of lingual and non-lingual segments on the rate of tongue movements in the AD group (see Hypothesis 9 in section 2.6), difference in the rate of the tongue movement between the AD and the TDC groups, and between the AD and the CAS groups (see Hypothesis 10 in section 2.6). In order to answer which method is better for the current data it is thus necessary to look at the pattern of results on individual tests in more detail.

One possible approach is to look first at the outcome of those hypotheses that are very strongly based on previous studies. For example, it has been shown in a number of studies that children have longer segment and syllable durations than adults and any other result would raise a certain amount of doubt. A method revealing such a difference can thus be expected to be more appropriate than a method not revealing the same. The two methods used here have shown exactly the opposite result on this comparison. The Mann-Whitney test showed no difference in syllable duration on any of the six syllables between the AD and the TDC groups. GLMM, on the other hand, revealed the expected result. Therefore, we can tentatively conclude that GLMM works better at revealing differences in syllable duration within and between the groups.

However, the question remains: how do we decide which is better in the case of the two methods showing exactly the opposite result on a new measure that has not been investigated earlier? We could examine the outcome for a hypothesis for which we very strongly expected a positive result. This was observed when testing Hypothesis 5 (see section 2.6) about the changes in the amount of tongue movement over a syllable caused by the addition of a lingual or a non-lingual segment. While the Wilcoxon Signed-Rank test showed that only half of tested syllable pairs in the AD and the TDC groups changed in the expected way, GLMM showed a significant effect of the number of lingual segments. Because of the way the amount of movement was measured in this study, it seems much more plausible that the addition of a lingual segment would increase the amount of tongue movement. A greater number of segments represents more position and shape targets that the tongue has to achieve in order to produce the correct output. For this reason, it seems that GLMM probably gives a more reliable answer than the non-parametric test.

Using both traditional non-parametric procedures and the novel method of applying GLMM to linguistic data, allowed a direct comparison of the methods and a better insight into the characteristics of the three speaker groups. All the answers to the

Hypothesis	Speaker group(s)	Non-parametric	GLMM
H1	AD	✓	✓
	TDC	✓	✓
	CAS	✓	✗
H2	AD: /p/	✗	✗
	AD: /s/	✗	✗
	AD: /l/	✗	✓
	TDC: /p/	✗	✓
	TDC: /s/	✗	✗
	TDC: /l/	✗	✓
	CAS: /p/	✓	✓
	CAS: /s/	✓	✓
	CAS: /l/	✓	✗
H3	AD vs. TDC	✗	✓
	AD vs. CAS	✓	✗
	TDC vs. CAS	✓	✓
H4	AD vs. TDC	✓✗	✓
	AD vs. CAS	✗	✗
	TDC vs. CAS	✗	✗
H5	AD	✓✗	✓
	TDC	✓✗	✓
	CAS	✓	✓
H6	AD vs. TDC	✗	✗
	AD vs. CAS	✓	✓
	TDC vs. CAS	✓	✓
H7	AD vs. TDC	✗	✗
	AD vs. CAS	✗	✗
	TDC vs. CAS	✗	✗
H9	AD: lingual	✓	✗
	AD: non-lingual	✓✗	✗
	TDC: lingual	✓✗	✗
	TDC: non-lingual	✗	✗
	CAS: lingual	✗	✗
	CAS: non-lingual	✗	✗
H10	AD vs. TDC	✓✗	✓
	AD vs. CAS	✓	✗
	TDC vs. CAS	✓	✓
H11	AD vs. TDC	✗	✗
	AD vs. CAS	✗	✗
	TDC vs. CAS	✗	✗

Table 5.1: The outcome of hypotheses testing by non-parametric and GLMM methods. Hypothesis was either confirmed (✓), not confirmed (✗) or the results do not allow a firm conclusion (✓✗).

hypotheses discussed in the following sections are based on GLMM.

5.2 Comparing the AD and TDC control groups

This section will discuss the results of the AD and the TDC groups by comparing them to each other and to previously reported speech characteristics relevant to this study. The decision to include two control groups was twofold: first, it allowed comparing the performance of speakers with CAS with mature characteristics of adult speech and with the speech of typically developing children who have not achieved adult-like speech control; second, inclusion of two previously well researched speaker groups served also as a verification system of the chosen research methods and methodological procedures.

5.2.1 Duration

Out of all the reported results, syllable and segment durations were the most traditional measures investigated in this study. The two control groups were compared to each other first with respect to syllable duration and within-group variability of syllable duration and second, in the way syllable and segment durations change as a result of changes in the syllable onset structure.

As expected, shorter syllable durations were found in the TDC groups than in the AD group, supporting Hypothesis 3 (see section 2.6), but surprisingly, the within-group variability followed the prediction (see Hypothesis 4 in section 2.6) only in three out of six target syllables. Longer durations of speech units (segments, syllables, words) and greater within-speaker and within-group variability in children have been found previously in a number of studies (Smith, 1978; Kent and Forner, 1980; Smith and Kenney, 1998; Lee et al., 1999) and have been explained as being the result of a still-developing speech control system. The discrepancy may be caused by the speaking style resulting from the characteristics of the speech material in the present study. All the target words and the following word ”today” ended with the same vowel, making it possible for the participants to slip into a rhythmic, poem-reciting, speech and producing all the utterances with similar durations. In that way target words tended to be more similar to each other and groups showed smaller within-group variability than expected. Such productions were observed for some of the TDC speakers but not for the AD ones who kept a conversation-like prosody throughout the recording. Different styles of uttering the speech material might thus explain the lack of difference in the variability of "say" and "slay" between the two groups and even greater variability of "lay" for the AD group. Unfortunately, by the time the effect of the speech material was noted, it was too late to change it because of the amount of data already recorded and annotated.

Another important factor necessary to achieve mature temporal structure of speech is appropriate adjustment of syllable and segment durations to the complexity of the speech material. The results showed that the TDC group, just like the AD group, increased syllable duration with the increase of the syllable onset segments, thus supporting Hypothesis 1 (see section 2.6). However, the control groups did not show the same patterns of adjusting segment durations as a result of changes in the number of syllable-onset segments (see Hypothesis 2 in section 2.6). In contrast to what was expected, the AD group adjusted only the duration of /l/ and made no adjustments to the duration of /p/ and /s/. The TDC group, on the other hand, significantly changed durations of both /p/ and /l/ in clusters as compared to the singletons.

Although the finding that the AD group adjusted the duration of only one segment seems unusual, it is in fact not different from previously reported findings. When looking at characteristics of segment durations of singletons and clusters reported by Haggard (1973); Klatt (1974); O'Shaughnessy (1974); Umeda (1977); Crystal and House (1988) it becomes clear that different patterns were observed in different studies. This can result from different speech material, as discussed in section 2.2.1.1, but it may also reflect characteristics of individual speakers. As argued by Levelt (1989), the syllable is the main unit of articulation. Its rate depends on the eigenfrequency of the moving mandible and tongue body involved in vowel production, it is intrinsically timed by carrying the duration parameters of a whole utterance, but it also has its own individual duration depending on its composition and the number of segments. This suggests that in terms of articulatory planning with the aim to produce acoustically correct output, temporal characteristics of syllables are more precise than those of segments. The finding that the AD group adjusted the duration of some segments but not of others in clusters as opposed to singleton would provide support for this idea. It may be that speakers have some level of freedom of adjusting the duration of individual syllable segments, as long as the total syllable duration remains appropriate.

The importance of the syllable as a motor planning unit can also explain the difference between the adult-like adjustments of syllable duration to the syllable structure and the non-adult-like adjustments of segment duration observed in the TDC group. The still-developing speech system of children can be expected to stabilise first those parameters that are more basic and contribute more to speech intelligibility. Only after the basic features are established, can the refinement of the system begin, by improving such characteristics as adjusting segment durations in clusters. For that reason, temporal features of syllables become adult-like before temporal features of segments.

5.2.2 Amount of tongue movement

Measuring the amount of tongue movement in the way described in this thesis was a novel approach to the study of tongue movements obtained by an ultrasound recording. Ultrasound data was previously used mainly to compare tongue contours at different time points but this thesis introduced the idea of quantifying the amount of tongue movements from a sequence of tongue contours. Including control groups of adults and typically developing children without speech impairments was particularly valuable in this case. The two control groups provided information on how the methodology works on typical populations.

As expected, the TDC group resembled the AD in the ability to increase the amount of tongue movement when a lingual segment is added to the syllable (see Hypothesis 4 in section 2.6). This finding supports the observation by Ostry et al. (1984) that children aged 6 years already adapt tongue dorsum displacement to the stress and voicing condition in an adult-like manner.

In contrast to the expectations (see Hypotheses 5 and 6 in section 2.6), the TDC group did not differ from the AD group in the amount of tongue movement on most syllables but only on two “say” and “lay”. Speakers in the TDC group showed a smaller amount of movement on both of them. Such result was rather surprising, especially in the light of some earlier studies and the fact that children have a smaller size of both oral cavity and the tongue. In an ultrasound study of tongue dorsum movement in /kaka/ and /gaga/ sequences, Ostry et al. (1984) observed that children produce less movement over the vowel than adults, achieving only about 2/3 of the adult values. The tongue dorsum displacement over the vowel represents the distance between the tongue dorsum at the velar closure to the lowest tongue dorsum position during the vowel production and back to the velar closure again. This observation can be directly related to the smaller amount of movement over ”say” and ”lay” observed in this study. Just as /k/ and /g/ are characterised only by the tongue movement in the dorsum part of the tongue, ”say” and ”lay” are primarily marked by movement in the front part of the tongue. If children in Ostry’s study produced 1/3 less movement in the dorsum region than the adults, it is possible that children participating in this study produced similarly less movement in the front part of the tongue. This difference in movement was big enough to influence the amount of movement over the entire syllable. But why then, did the TDC group not produce less movement over the rest of the syllables? In the case of ”pay”, the measure could be affected by the lack of a lingual consonant. Since /p/ does not involve a lingual gesture it did not contribute to the movement and the amount of movement was in reality measured only from the position of preceding /ə/ over the vowel /e/ and this amount seems to be the same as in the AD group, showing no effect of the size of the speaker's tongue and oral cavity on the measure. In the case of the clustered onsets, however, another important factor must be considered. As most recently shown in an EPG study by Cheng et al. (2007a) children have a tendency to move their tongues as one articulator over the cluster /kl/ and lack adult ability to separate tongue tip and tongue body movements. Although both /s/ and /l/ have a closure at the same place, they differ in the shape the tongue has to form to produce the correct output. Tongue movements over /s/ and /l/ in clusters including both of these segments can be greater when compared to adults because of the whole tongue being involved in their production and not only the part necessary to produce the sound. In that way the TDC speakers would produce more movement and with this compensating for the otherwise lesser movement in each individual part of the tongue, causing the total amount of movement over a syllable to be no different from the AD group. Following this path of thought it would be necessary that the amount of movement over ”play” was the same as over ”lay” and thus smaller in the TDC group than in the AD. This is, however, not the case. The beginning and the end of a syllable were selected on the basis of the acoustic signal and tongue movements were measured over this selected part. In the case of “lay”, the tongue is already at the alveolar ridge at the beginning of /l/, but in the case of “play” the tongue is not involved in the production of bilabial /p/ and starts to move from /ə/ to /l/ after the beginning of /p/. More movement is therefore visible over “play” than over “lay”.

On the other hand, it is also possible that the results truly represent the characteristics of tongue movement in the AD and the TDC groups. This study is the first in which tongue movements were measured over a syllable from a sequence of mid-sagittal tongue contours. Previous studies dealt mainly with either movement in the individual parts of the tongue or the formants extracted from the acoustic signal. It is possible that investigating those measures revealed differences between adults and children but looking at the whole tongue surface does not. Since all speakers in the

AD and the TDC groups produced all perceptually correct segments and none of them had any speech impairment they had to achieve similar tongue shapes and positions which could be reflected in a similar amount of tongue movement. The measure was smaller for the TDC group than for the AD one only on the two syllables with most prominent movement in only one part of the tongue which could made the individual characteristics stand out more since measuring the amount of tongue movement over "say" and "lay" is more influenced by movement of an individual part of the tongue. Additionally, just as in the case of adjusting segment duration in syllable onset clusters, speakers might have a certain degree of freedom on how to move their tongues as long as they achieve the correct position and shape necessary for the correct production. This has been suggested earlier in the EPG study by Cheng et al. (2007a). They showed that it is not necessary to produce only one, typical, tongue-to-palate contact pattern in order to produce correct speech sounds but that a number of patterns are possible. Interestingly, teenagers and adults participating in that study produced less segments with the typical EPG pattern than the younger two groups. Although the present study significantly differs from the above one in the method used to investigate tongue movements, the results are still rather compatible.

The same reason might also explain the lack of differences in the within-group variability. Although greater variability of tongue movements was reported earlier in acoustic (Kent, 1976; Nittrouer, 1993; Smith and Kenney, 1998) studies it has to be pointed out that findings of acoustic studies raise some doubt due to the uncertainty about measuring formant frequencies of child speech (Lindblom 1972, cited in Kent 1976, p.434). Articulatory studies of tongue movements in children provided a rather different conclusions. Although age was found to have a significant effect on the variability in the placement of tongue-to-palate contact in /k/, and that 6-7 year-olds were more variable in the contact between tongue body and palate than older children and adults (Cheng et al., 2007b), there was no difference in the within-speaker variability in the duration of the approach, closure or release phase between any of the investigated age groups (Cheng et al., 2007a). The fact that the TDC and the AD groups did not differ in the within-group variability might result from allowed range of tongue positions and shapes that still produce the correct output (Cheng et al., 2007b,a). Younger speakers in the TDC group can be expected to have less developed motor control system but because the range of possible tongue positions and shapes might be relatively large, they stay within it and not differ in the amount of variability. Additionally, as demonstrated by Lindblom (1990), unconstrained movements tend to operate at a low-cost form of behaviour meaning that tongue movements would be always realised in the most economical way that still lead to the correct output. The usage of the greater range of articulatory options in adults could be motivated by this economy criteria.

Such an interpretation would lead to a conclusion that the TDC children aged six to nine years already possess adult-like control over global tongue movements although they still lack control over fine motor movements of individual parts of the tongue (as shown in previous studies mentioned above). However, the allowed range of variation in tongue positions allows them to produce correct output in spite of still immature motor control abilities. Additionally, the results showed that measuring the amount of movement in the way described in this thesis is a useful measuring technique since it presents the movement of the entire midsagittal tongue contour but the interpretation of the data also benefits from information obtained in studying movement in individual parts of the tongue.

5.2.3 Combining temporal and articulatory data

A great advantage of this study was that acoustic and articulatory data could be used simultaneously in an analysis of the same spoken output, enabling access to a much richer source of information.

The first attempt at combining the two types of data was made by calculating the rate of tongue movement over the syllable. As expected (see Hypothesis 10 in section 2.6), speakers in the AD group achieved significantly greater rates than the TDC group, reflecting differences in the syllable duration and the amount of tongue movement, and displaying the functioning of a developed speech motor system that can achieve goals fast and efficiently. But again, the two control groups did not differ in the amount of within-group variability (see Hypothesis 11 in section 2.6).

Although hypothesised (see Hypothesis 9 in section 2.6), the combined data revealed that none of the speaker groups showed any effect of adding either lingual or non-lingual segments on the rates of tongue movement. Such a result is rather surprising because the AD and the TDC groups showed a significant effect of number of onset segments on syllable duration, and they both increased the amount of tongue movement with the addition of a lingual segment. However, it seems that these effects were not big enough to also show the combined effect on the rate of tongue movement.

In addition to calculating the rate of tongue movement, acoustic and articulatory data were used to inspect the distribution of the syllables by plotting the amount of tongue movement over a syllable as a function of syllable duration and colour-coding the types of syllables. One of the ideas on how to use both types of data at the same time was to use a classification system for assigning data points into different categories and to check whether the data from different speaker groups ended in different categories. After a first inspection of the measurements it became obvious that there was a great overlap between different syllables inside groups and also between groups. The classification was not further employed as it became clear that it would not reveal any useful information. However, just analysing the combined plots showed a difference between the control groups, with the TDC group having a greater spread of values in all syllable categories than the AD group. Such observation most likely reflects poorer control of speech production, both in acoustic and articulatory domain, in children.

5.2.4 Tongue movement patterns

Besides quantifying the amount of tongue movement, ultrasound recording also enabled observation of tongue movement patterns over a syllable by imaging the whole midsagittal tongue surface during speech. Observing these patterns was expected to reveal important differences between the speaker groups (see Hypothesis 8 in section 2.6).

As predicted, the AD and the TDC speakers produced similar patterns over the same syllables but were different in the stability of the patterns over the repetitions. While the AD speakers produced very similar patterns in all five repetitions, the TDC speakers performed differently. They all showed two or even three distinctive patterns in the five repetitions of the same syllable. Interestingly, the same was observed when looking at the patterns formed by a sequence of entire midsagittal tongue contours or only of the highest points on these contours. Such observation was not surprising since the TDC speakers are still acquiring mature speech motor abilities and is in line with previous observations by Cheng et al. (2007c) and Goozee et al. (2007) who observed that children up to 11 years of age display different patterns of tongue movement than adults. The findings additionally confirmed the power of ultrasound data to bring out differences in tongue movement between different speakers.

Another attempt at analysing tongue movement patterns was made by observing whether the midsagittal contours form an arch or a pivot pattern in transition from one segment to the other (see description of the study by Iskarous (2005) in section 2.3.1).

Unfortunately, this method did not distinguish between the AD and the TDC speakers. Overall, all speakers produced very similar transition patterns but not even the three AD speakers produced exactly the same patterns. It is important to stress that all AD and the TDC speakers produced perceptually correct segments and the slight differences in the patterns probably resulted from individual differences in articulation characteristics. As mentioned earlier, segments are not articulated in only one possible way and a speaker may have a certain amount of freedom when planning and articulating tongue movement sequences. This may lead to a relatively high variability when data of only few speakers are analysed.

5.2.5 Summary

In summary, the AD and the TDC groups showed the expected results and the relationships between them on some measures but not on others. Because of their mature speech production system, the AD group produced longer and less varied syllable duration than the TDC one. Both groups adjusted syllable durations to the number of syllable-onset segments but they differed in the adjustment of segment durations in clusters as compared to the singletons. The two groups did not differ in the amount of tongue movement over syllables and they both adjusted the amount of movement to the number of lingual syllable-onset segments. No difference was observed in the control of rate of tongue movement over the syllables. Simultaneous inspection of temporal and articulatory data of individual syllable types additionally revealed that children of this age produced a greater range of values than adults, again showing poorer control of both domains. Although the groups could not be distinguished on all the articulatory measures, ultrasound data became very useful when observing tongue patterns and consistency of patterns in repetitions. Here, the TDC group revealed that at 6 to 9 years of age, speakers can already produce adult tongue movement patterns over syllables but they are still less consistent in their repetitions.

5.3 Comparing speakers with CAS to the AD and the TDC groups

This section will discuss results of the CAS group and individual speakers by comparing them to the control groups presented in the previous section, and to the results of earlier studies of CAS and the speech characteristics used in a diagnostic procedure.

5.3.1 Duration

Because one of the commonly cited characteristics of CAS is inconsistency of production (Davis et al., 1998; Maassen, 2002; Forrest, 2003; Marquardt et al., 2004; Betz and Stoel-Gammon, 2005; Davis et al., 2005; Peter and Stoel-Gammon, 2005; American Speech-Language-Hearing Association, 2007; Flipsen Jr. and Gildersleeve Neumann, 2009), it was expected that speakers with CAS would produce some syllables with durations similar to the AD and the TDC speakers, some longer and some shorter, and as a group, would not differ in syllable duration from the control groups but would show greater within-group variability on this measure. This turned out to be only partly true. The CAS group produced longer syllable durations than the AD group on four out of six syllables but at the same time did not differ significantly from the TDC group. More interestingly, the CAS group showed the same amount of within-group variability as the AD and the TDC groups. These results thus only partly support Hypothesis 3 and do not support Hypothesis 4 (see section 2.6). They are also in contrast to previous studies suggesting longer syllable durations in CAS. Although not directly reporting syllable durations, Nijland et al. (2003b) observed that children with CAS produced longer durations of all segments in tested syllables than control children and Peter and Stoel-Gammon (2005) reported that they produced longer vowels and similar onsets in monosyllabic words than control children. Both these results suggest greater total syllable duration. Less agreement was achieved previously on the topic of within-group variability. It has been reported that groups of speakers with CAS were more variable on a wide variety of timing tasks (Peter and Stoel-Gammon, 2005, 2008) and in all segment durations which were part of a syllable (Nijland et al., 2003b), but had a lower coefficient of variation on speech events (Shriberg et al., 2003a) than a group of control children. None of the studies, however, reported the same amount of group variability in CAS as in the control groups.

Why does the CAS group in this study then differ from adults but not from children in syllable duration and not differ from any of the control groups in within-group variability? First, the characteristics of syllable duration could be attributed directly to the speech impairment. Participants with CAS are three and eight years older than the oldest children in the TDC group. Without the speech impairment they would perform on the adult level and not resemble young children any more. But CAS, with an impairment at the speech planning level, is expected to influence all levels of speech planning, including temporal structure and for that reason it is not surprising that the CAS group performed immaturely and more similarly to the TDC than the AD one on the measure of syllable duration. However, because of their age, they might have different syllable durations than previously reported. Participants in previous studies were always younger children with CAS and not teenagers. Second, the reason for the observed lack of higher within-group variability may be the small number of participants which is unfortunately a persistent problem of clinical studies. A greater number of speakers would represent group variability more accurately as it would be less affected by each individual speaker's performance. It is, however, also possible that the three speakers with CAS performed more consistently than reported previously as being typical for CAS and in that way showed similar within-group variability as the control groups.

Investigating the effect of syllable structure on syllable and segment durations revealed even more similarities between the CAS group and the control groups. The finding that the CAS group increased syllable duration with the addition of a syllable-onset segment, like the control group, does not support Hypothesis 1 (see section 2.6). The outcome of Hypothesis 2, on the other hand, is more difficult to answer. It was confirmed to the extent that the CAS group did not show the changes in segment durations that were expected in the AD and the TDC groups, but it was not confirmed that the CAS group would perform differently than the control groups. Just as the AD group did, the CAS group adjusted only the duration of /l/ and in this respect their performance was more adult-like than the TDC one.

Because there were only three participants with CAS it was necessary to evaluate not only the group results but also results of individual speakers. This approach seemed to be valuable as it showed that none of the speakers with CAS produced longer or shorter durations than any of the AD and the TDC speakers. CAS1 produced the most adult-like durations, while CAS2 and CAS3 were more similar to the TDC speakers. Interestingly, their ranges of measured durations were comparable to those of the AD speakers on most of the syllables and smaller than in the TDC group. Such an observation is in contrast to previously reported general inconsistency of speech production in CAS. If that had been the case, speakers with CAS should display grater ranges than at least the AD speakers. Unfortunately it is not possible to conclude whether this observation is a true representation of CAS (with all three speakers just happening to produce consistent duration in the recording session), the effect of their age or speech therapy. The question thus remains open and it calls for further research.

All three speakers with CAS also showed sensitivity to the number of syllable segments and adjusted syllable durations accordingly. Two of the speakers (CAS1 and

CAS3) additionally showed the adult-like ability to adjust onset-segment durations according to different numbers of onset segments. Both of them and the AD group significantly adjusted only the duration of /l/ while speaker CAS2 showed no adjustments of segment durations in different onsets. Only speaker CAS2 thus performed similarly to previously reported observation that speakers with CAS do not change segment durations in clusters (Nijland et al., 2003b). However, the participants in that study were much younger (4 to 6 years, compared to 14 and 18 years in the present study) and for that reason the observed differences are not necessarily surprising. Results of the study presented here suggest that some older speakers with CAS overcome a speech characteristic observed in younger children with the impairment and achieve mature adaptation of segment duration in different contexts. But at the same time it can be concluded that the same characteristic can be very persistent, as CAS2 still showed different production at the age of 18 years. The adult-like level of duration control observed in the other two speakers with CAS could be attributed to either maturation of speech production system with age or to speech therapy. However this is difficult to argue in the light of speakers CAS2 and CAS3 being identical twins, growing up together and probably receiving similar amounts of therapy.

Based on these results it would seem that speakers with CAS demonstrate almost mature temporal speech characteristics with adult-like ranges of measured values and adaptation of duration in different syllables. On the other hand, the results suggest that adjustment of segment durations becomes adult-like for some speakers with CAS but not for all of them. Measuring syllable duration also revealed that this measure alone is not enough to separate speakers with CAS from either of the control groups since, although as a group they showed some longer syllable durations, individually they did not produce higher or lower values.

5.3.2 Amount of tongue movement

The amount of tongue movement was measured directly from the ultrasound imaging of the oral cavity and was the central measure of this study. Because of the nature of CAS, it was expected that this measure would reveal significant differences in tongue movement control, demonstrated particularly in greater variability and different adaptation to the increased demand of motor planning imposed by consonantal clusters as opposed to singletons.

When comparing the amount of tongue movement and within-group variability over the syllables it became clear that the CAS group did not differ from either the AD or the TDC groups, with the exception of more movement over "lay" than the latter. This finding supported Hypothesis 6 about the CAS group not having a significantly different amount of tongue movement over syllables than either of the control groups, but not Hypothesis 7 about greater within-group variability in the CAS group than in the control ones (see section 2.6).

The finding that the CAS group is not significantly different from the control groups on the measure of the amount of tongue movement was expected and is believed to arise from the characteristics of speech impairment. Without the speech impairment teenagers would be expected to resemble more the AD group than the much younger TDC group. But because of their problem with spatiotemporal planing and resulting inconsistency of production, they were expected to show average values that fall between those of the TDC and the AD groups. Further investigation of individual speakers with CAS supported this, with speakers with CAS showing the same relationship to speakers in both control groups, but also revealed differences between the three speakers. Speaker CAS1 was the only one achieving smaller values and in that way showing less movement over some of the syllables than control speakers. In contrast, speaker CAS3 showed a greater amount of tongue movement over some syllables than either AD or TDC speakers. Speaker CAS2 moved his tongue a similar amount as most of the control speakers. The three speakers did not show the same ranges of values on each of the six syllables but it could still be concluded that their ranges were comparable to or even smaller than those of the AD and the TDC speakers, again showing rather consistent production.

Following the observation about ranges of measured values, it is not so surprising that within-group variability was not greater in the CAS group than in the control ones. The same observation about the within-group variability was reported previously by Nijland et al. (2002) but because it was based only on F2, it was expected that measuring the amount of entire tongue movement would have greater power to depict the influence of the impairment and thus show greater variability. Although the measure of within-group variability of the amount of tongue movement in this study is affected by a small number of participants, the fact that all three speakers showed a rather narrow range of measured values supports the validity of the results. Of course, more participants would be needed to answer the question in greater detail.

Another surprising result was obtained when investigating the influence of the type of syllable-onset segments on the amount of tongue movement. Planning of movement sequences was expected to be more challenging when articulating clusters with two lingual segments than singletons or clusters with one lingual and one non-lingual segment. The results showed that the CAS group reacted the same as the control groups, by increasing the amount of tongue movement over syllables with a greater number of lingual segments (see Hypothesis 5 in section 2.6). However, when looking at the individual speakers with CAS, two out of three showed the same effect and one, speaker CAS2, did not.

What can be concluded about the amount of tongue movement over syllables in CAS? First, the measure showed that speakers with CAS showed somewhat different characteristics. Two of the speakers performed at the extreme ends of the control groups (being similar to the speakers with the most or the least tongue movement) but reacted the same to the type of syllable-onset segments, while one speaker showed a similar amount of tongue movement but no effect of the type of syllable-onset segments. These kinds of differences are likely to be one of the main reasons that the impairment remained controversial for a relatively long time and that its nature and characteristics are still not well understood. This study is a contribution to the understanding of speech motor control in CAS and although it cannot fully answer the question about the characteristics of the amount of tongue movement in CAS, it does contribute to the understanding of the problem by showing that although as a group speakers with CAS do not differ from adults and typically developing children, individually they show different characteristics.

5.3.3 Combining temporal and articulatory data

Because of the impaired timing and motor control in CAS, it was expected that combining temporal and articulatory data in the form of rate of tongue movement would uncover great differences between the CAS and the control groups. However, this was not entirely true.

The CAS group had slower rates of tongue movement than the AD group, but with the exception of faster rates over ”say”, it did not significantly differ from the TDC just as these two groups did not differ in the syllable durations or in the amount of tongue movement over the syllable (see Hypothesis 10 in section 2.6). The characteristics of these two measures were reflected in rate of tongue movement of individual speakers with CAS as well. Speaker CAS1 produced the slowest and the most child-like rates of tongue movement while speaker CAS3 was the most adult-like. Furthermore, their ranges of values again did not differ from ranges observed in individual control speakers.

The three groups additionally did not differ in within-group variability (see Hypothesis 11 in section 2.6) which was not surprising since almost no difference was observed on the within-group variability of the other two measures. It was however surprising that, like the control groups, the CAS group and individual speakers did not change the rate of movement as a result of changes in the type of syllable onset segments (see Hypothesis 9 in section 2.6).

But plotting duration and amount of tongue movement by syllable type and speaker group did show some interesting differences between the groups. Of course, it has to be stressed that the group results of the CAS speakers are affected by roughly two thirds less data than the other two groups and for that reason show less variability. However, some conclusions about the impairment can still be made. As presented in Chapter 4, the CAS group and individual speakers showed similar relationship between syllable types as the control groups, with the spread of syllable categories being more similar to the AD group than to the TDC one. At the same time the CAS group also showed a narrower spread of all measured values than either of the control groups, again supporting the conclusion that speakers with CAS do not produce wider ranges of values in either the temporal or articulatory domain and that the consistency of their productions was not reduced.

These results lead to a conclusion that combined temporal and articulatory data provides a more informative view of speakers' performance and can help in revealing speech characteristics of CAS by differentiating them from the control groups. Research of CAS can benefit from joint analysis of both of these domains.

5.3.4 Tongue movement patterns

The possibility of obtaining tongue movement patterns was one of the main motivations for this study. It was believed that observing tongue movement patterns could reveal important characteristics of CAS since what better way is there to investigate motor speech impairment than by directly observing articulators during speech?

The midsagittal tongue contour patterns revealed that the speakers with CAS had a lower position of the front of the tongue and a higher middle and even back part of the tongue during the production of a vowel than the control groups. They all showed similar tongue movement over the single-segment /p/ and /s/ onsets to the AD and the

TDC speakers but different patterns over singleton /l/ and all clustered onsets. This difference can be attributed to the level of articulatory demand imposed by a particular segment or combination of segments. /p/ as a non-lingual segment does not contribute to lingual movement and does not increase the information load on the motor system planning of tongue movements. /s/, on the other hand is a lingual segment but as part of a sequence ”a say” does not contribute to great changes in the tongue position or shape since movement happens only in the front part of the tongue with the back and even the middle staying stable during the articulation of /s/ in this sequence. The lack of /s/'s power to distinguish between speakers groups, particularly the CAS speakers, may additionally lie in the main limitation of ultrasound recording that raised tongue tip cannot be imaged. It is possible that the tongue tip was not visible on the scanned images and its movement could not be observed. If that information was present, the three groups may have been more likely to show different patterns over ”say” as well. However, the CAS speakers could be distinguished from the control groups when comparing tongue movement patterns over onsets that included /l/. In order to articulate /l/ in the sequences used in this study, all parts of the tongue had to be involved. The back of the tongue was raised higher than the front which had to move upward into the following vowel. Such a tongue configuration is highly demanding and impairment that affects tongue movement planning is likely to be reflected in the articulation of /l/. Additionally, syllables including /l/ also revealed some differences between the speakers with CAS. While CAS1 and CAS2 both lacked raising of the back of the tongue which was lower than the raised front part, speaker CAS3 achieved a similar tongue configuration to that of the AD and the TDC speakers. More difference between speakers with CAS was observed also when comparing tongue movements over syllables with clustered onsets only. The patterns of speakers CAS2 and CAS3 showed an effect of onset structure with similar patterns over onsets with two lingual segments (”slay” and ”splay”) but a different pattern over onsets with a non-lingual/lingual combination (”play”). Speaker CAS1, on the other hand, did not show the same effect and all his patterns over clustered onsets looked similar.

Another important aspect of tongue movement pattern is their similarity over repetitions. Out of three analysed speakers, only CAS3 produced different midsagittal tongue contour patterns in the repetitions of only two of the six syllables. When looking at patterns formed by the highest point on the tongue contours, speakers with CAS showed as stable repetitions as the AD speakers and more stable than the TDC speakers. Such a result was rather surprising because the CAS group was expected to show the most variation in tongue movement patterns due to their impaired speech motor planning and high inconsistency in speech production.

Taken together, it can be concluded that speakers with CAS have different tongue movement patterns than the AD and the TDC speakers but they do not differ in the variability of production from the mature AD speakers. The experiment also demonstrated that although observing the entire tongue contour to explore speech characteristics of CAS is important, it would be additionally necessary to observe movement of separate parts of the tongue and coordination between them.

The analysis of tongue movement patterns showed that CAS does affect tongue movement and that speakers with CAS produce qualitatively different patterns. This is especially interesting in the light of the fact that all speakers with CAS produced perceptually correct speech segments. Although not directly experimentally evaluated, observing ultrasound video recordings gave the impression that speakers with CAS moved their tongues less and used their oral space in a different way during speech than the AD or the TDC speakers, especially in the vertical direction. Although all speakers articulated correct targets they differed in the extent of the realised articulatory movements. An individual segment is not always articulated at exactly the same place in the oral space but it can slightly vary around the most optimal location. The same acoustic properties of a speech sound can be achieved by positioning the tongue inside a target area and not at only one particular place. Articulatory movements are additionally influenced by economy criteria (Lindblom, 1990). Both of these characteristics of speech production are especially evident in the case of coarticulation when a balance between the optimal realisation of each segment and ease of articulation have to be achieved. It was reported earlier (Nijland et al., 2002) that children with CAS coarticulate more than typically developing children and more than adults. That could mean that they less often realise articulatory movements in their optimal form and more often in a reduced one. They move their tongues just enough to reach the area and position which results in the correct acoustics of the spoken output. Such behaviour could be a result of the impaired motor planning system simplifying the demands brought upon it in the attempt to produce correct output. However, because the productions are always on the edge of the area that still results in the correct output, it is easy to slip over the edge and out of the target segment area. This can either result directly in the production of a wrong segment, if the tongue configuration happens to be compatible with another speech sound, or, through feedback, in an attempt to correct the production, which causes a break in the planned articulatory sequence and re-planning. Such an attempt at an on-line correction is likely to be too demanding for an impaired speech production system and the resulting output is wrong and intelligibility of speech reduced. Both a high number of segmental errors, vowels and consonants, and groping movements are typically associated with CAS. Although they were not observed in the speech of participating speakers with CAS, the obtained data provides a potential explanation for their occurrence as well. Additionally, the observation that speakers with CAS use less of their oral space also explains differences in their midsagittal tongue contour patterns over the syllables and similar consistency of movement patterns in the repetitions as the AD speakers. If they use less space they also have less chance for higher variability.

The observation of less tongue movement in CAS was also discussed at the 2009 ASHA convention with several speech and language therapists who work with speakers with CAS and some of them had the same impression about their clients. In the light of their replies and the observations of this study it seems valid to continue investigating tongue movements in the speech of speakers with CAS as the information could contribute significantly to the ease of diagnosis and planning of therapy procedures.

5.3.5 Summary

Overall, this study showed that speakers with CAS cannot be easily differentiated from adults and typically developing children solely on the measures of syllable duration, amount and rate of tongue movement, or in the control of these measures. Additionally it revealed that group results of speakers with CAS do not always represent the characteristics of individual speakers.

The results showed that although as a group they produced longer syllable durations, slower rates of tongue movement and similar amounts of tongue movement compared to the AD group, only speaker CAS1 showed values on individual measures that fell out of the range of the AD and the TDC speakers (lower amount of movement on most syllables and lower rate of movement on some syllables). The observation that speakers with CAS achieve similar values on a number of measures can explain the fact that CAS remained poorly understood for a relatively long time. If speakers displayed significantly different results and produced values outside of the range of control groups, it would be much easier to set diagnostic criteria.

Similarly, speakers with CAS showed adult-like ability to adjust syllable duration to the number of onset segments with two of them additionally performing at adult level in adjusting amount of movement, rate of movement and segment durations to a specific syllable structure. Some speakers with CAS are thus able to develop mature control over some aspects of speech production by their teenage years, while others (CAS2 in this case) show deviant performance on all but the most basic feature of syllable duration.

Although the CAS group did not significantly differ from the TDC group on syllable duration or the amount of movement, the two groups (and individual speakers), differed in the range of measured values. As a group, CAS speakers showed a narrower spread of individual syllable categories and of all data points than the TDC groups. This was further reflected in individual speakers with CAS who produced narrower or similar ranges of measured values on most syllables. Such observation is in contrast to commonly reported high inconsistency of different aspects of speech production in CAS. None of the measures addressed in this study supports it. High consistency was additionally observed in the patterns of tongue movement over syllables. Although speakers with CAS produced different patterns than either of the control groups, they were very consistent in their repetitions. Of course, high consistency could result from low number of participants and/or repetitions and more data is needed to reveal how (in)consistent on acoustic and articulatory measures speakers with CAS truly are.

Finally, the study showed that CAS is truly a rather complex disorder and that revealing its speech characteristics can benefit from acoustic and articulatory analysis of the same data. It also highlighted a valuable contribution of ultrasound imaging since observing tongue movements during speech revealed different tongue movement patterns in CAS and suggested that speakers with CAS use less of their oral space.

5.4 Methodological issues

Before finishing the discussion of the results obtained in this research, I would like to point out some other possible sources of influence, such as issues related to participant recruitment, ultrasound imaging, selection of speech material and measurement technique.

One of the major problems of studies of speech impairments is a low number of participants. This influences the conclusions that can be made about the speech impairment as they can be affected by high between-speaker variability. At the same time it reduces the validity of conclusions about group variability when comparing groups with speech impairments to the control group. Research into CAS is no exception to this. In fact, it its even further complicated by the still uncertain set of diagnostic criteria for the impairment.

Ideally, the participants would be young children, recorded before they start with speech therapy. In this way it would be possible to collect "pure" apraxic speech and potentially reveal a set of main speech characteristics. However, selecting young children can be rather controversial because their diagnosis is not based on firm criteria and because of the difficulty distinguishing between motor and phonological speech disorder at a young age. This can be additionally made difficult by child speech being very unintelligible or practically absent.

In the case of articulatory studies using ultrasound (but also EPG and EMA), young children are additionally not suitable because of the requirements of the recording procedure. During an ultrasound recording, a participant has to sit still to achieve the same scanning view throughout the recording session. Young children might not be able to achieve that. For the same reason, the ultrasound probe has to be fixed under the participant's chin and in the case of the headset used in this study this would not be possible for young children because of the weight of the headset. Additionally, young children would be very likely to produce more segmental errors, either as a result of the typical speech acquisition process or the impairment. The produced segments would be different between speakers, making it impossible to compare tongue movements. The obtained measures would not show the effect of the impairment but of articulated segments.

For these reasons, it seemed more appropriate to recruit older children and teenagers. First of all, because they are able to participate in an ultrasound recording, but also because they were most likely diagnosed in early childhood and the fact that they are still diagnosed with CAS means that they probably truly have it and not some other kind of speech impairment. However, this brings in the additional problem of the influence of speech therapy. Teenagers with CAS have usually been involved in speech therapy for a number of years and even though their speech has not improved enough to be declared as not having CAS any more, they might show some characteristics of over-trained speech elements. For example, if a speaker had more problems articulating a certain consonantal cluster, that cluster could have been trained to the point that on demand, such as in an experimental condition, the speaker does not show any problems with it although in spontaneous speech the problem might still be present. Similarly, over-training can be observed in controlling temporal or prosodic features of speech. Without an extensive review of a speaker's speech therapy records, interviewing the speech and language therapists working with the participants, and the application of a number of speech tests it is difficult to say whether a particular speech characteristic is over-trained or whether it truly represents the participant's speech. Because such extensive investigation is usually not possible, speech is described as it is but I think it is important to keep in mind that some detailed features could be influenced by therapy.

Regarding ultrasound imaging, I would suggest one change if this study was to be replicated. As a standard part of the ultrasound recording, the participants should have their palates traced. This is a relatively straightforward procedure in which the participants are asked to swallow a little water. Because the density of the tongue tissue and water are quite similar, the ultrasound waves do not get reflected at the tongue-water boundary but only at the water-palate one. This results in an outline of the hard palate. The information is necessary for the evaluation of the shape and size of the speaker's oral space and it would make it possible to better quantify how much of the oral space speakers use. This would allow investigating in greater detail the observation about speakers with CAS using less of their oral space than the control speakers.

Another methodological issue that has to be pointed out is that the results reported here describe the characteristics of tongue movement only in a single vowel environment /e/. In order to fully investigate the effect of clusters on tongue movement it would be necessary to replicate the study using another vowel in the speech material. The motivation for this is three-fold.

First, it has been reported that speakers with CAS coarticulate more than TDC, who in contrast coarticulate more than adults (Nijland et al., 2002). Coarticulation is additionally extended more forward in the CAS group (Maassen et al., 2001; Nijland et al., 2003b) and for this reason it is possible that the chosen vowel affects the entire cluster, makes it more similar to the particular vowel and by doing that, potentially makes all onsets followed by the same vowel more similar to each other. Including a different vowel would allow inspecting the exact influence of vowel type on the tongue movement patterns over clusters and also extracting tongue movements over clusters by speculating about the elimination of the vowel effect.

The second reason for including another vowel is that one of the commonly reported characteristics of CAS is frequent vowel errors

The final reason for including another vowel is the elimination of the rhyming effect in the speech material used. Although this could be also achieved by changing only the word following the target words, the rhythmic pattern of uttering one prompt after the other could remain. However, having two sets of speech material differing only in the vowel type would reduce the rhythmic pattern of the recording session as the prompts with different vowels could be alternated.

Inclusion of another vowel in the target words could lead to at least two interesting research questions. How does the type of vowel affect tongue positions and movement patterns in preceding singletons and clusters? Do speakers with CAS show impaired tongue movement in the articulation of vowels?

Finally, because the technique for measuring the amount of tongue movement over syllables was a novel approach to investigating tongue movements in continuous speech, it calls for some additional comments as well. Because the measure of amount of tongue movement is based on the average NND between consecutive pairs of tongue contours it does not provide any information about the different amount and direction of movement in different parts of the tongue. As discussed earlier, such a measure is still informative but it could benefit from further application to different speech material in order to test the extent of its validity. If it was measured over all possible segments, clusters and/or syllables in a given language, it would be possible to extract first those segments or combinations of segments that influence it the most and those that do not. This information would further allow extracting those articulatory features that affect such a global measure of tongue movement the most, and it would also reveal the influence of the amount and the direction of movement in different parts of the tongue. Investigating the amount of tongue movement over such diverse speech material would additionally make it possible to relate it and the findings of what kind of articulatory movements affect it most, to the perceptual and acoustic characteristics of the spoken output. Such a complex process would provide necessary information about the power of the measure to distinguish between small differences in the investigated speech material and about its validity.

5.5 Future research

As mentioned in Chapter 2.1, CAS is still not well understood speech impairment. Its existence has only been firmly acknowledged since 2007 after ASHA published a technical report on the disorder. That paper (American Speech-Language-Hearing Association, 2007) defined CAS as a disorder with problems in planning or programming speech movements in the absence of neuromuscular deficits, but also showed that the cause, prevalence, influence of genetics and co-existence with other neurobehavioral disorders are still poorly understood or even unknown. Additionally, it was stressed that CAS can include a number of speech characteristics which can be observed in other speech disorders as well. Differential signs of CAS were presented only very recently at the ASHA convention in 2009 (Flipsen Jr. and Gildersleeve Neumann, 2009). They included inconsistent output on repeated attempts of the same words, disrupted and lengthened transitions, and disordered prosody.

The first two of these characteristic are very suitable for articulatory analysis. Articulatory methods seem appropriate to investigate CAS because they can provide a richer source of information not only about articulatory movements but also about temporal and prosodic properties of speech resulting from a disruption of motor programming. Out of the three available methods for observing tongue movements, EPG, EMA and ultrasound, the latter is the least invasive and for that reason probably the most appropriate for the collection of CAS speech data. Additionally, investigation of both inconsistency in repetitions and disrupted transitions could be based on the research and results presented in this thesis.

The first measure worth further development is the amount of tongue movement over a sequence of segments. Measuring the amount of tongue movement across syllables as presented in this thesis did not reveal any important differences between the CAS and control speaker groups, suggesting no difference in overall tongue movement. However, this measure incorporated only the average distance between two tongue contours without any acknowledgements of the direction of the movement or the part of the tongue involved in the movement. Following the observation that speakers achieve better movement control of different parts of tongue with age and maturation (Cheng et al., 2007a; Goozee et al., 2007), it might be beneficial to measure movement of different parts of tongue in CAS as well. Such measures would not only provide information about whether speakers with CAS control their tongues more like young children, more like adults or in their own way, but they could also be incorporated into a new measure of overall tongue movement. Exact division of the tongue into separate parts is unfortunately not straight-forward because there are no stationary points on the tongue that could serve as a land-mark. However, if not too many parts are needed (e.g., only front and back region of the tongue), this could be achieved by producing an image of the tongue in a neutral position during rest, using the division of the oral space to separate the tongue into parts and then use that information to decide on the tongue parts in tongue contours extracted from continuous speech. Additionally, a faster frame rate creates more images per second, resulting in shorter time difference between two images and thus making tracking of changes in two consecutive tongue contours easier and more reliable.

In order to be able to divide oral space into several regions, all participants should have their palates traced during the ultrasound recording. Together with the acoustic shadows resulting from the mandible and hyoid bone, the information about the palate would completely enclose the oral space. This would additionally provide information about the oral cavity size and palate shape which would be beneficial for between-speaker comparisons.

A reliably marked oral space would allow further exploration of the hypotheses presented in this thesis, namely that speakers with CAS use less of their oral space than the control speakers, especially when articulating consonantal clusters. It would be interesting to investigate whether this is due to greater coarticulation, as observed by Nijland et al. (2002), or results directly from the impairment and just gives the impression of greater coarticulation. However, if it was confirmed that speakers with CAS use less of their oral space during speech, this could be incorporated into diagnostic procedures as it can be potentially easily verified by ultrasound imaging.

In order to support the above hypothesis, more data from different speakers with CAS, speakers with similar speech disorders and speakers with typical speech development are needed. Investigated speech material would also have to be expanded to include more and different consonantal clusters and different vowels as discussed in Section 5.4.

Up until now, ultrasound usage was a bit limited because of a lower rate of creating imaged frames and because tracing the tongue surface was only partly automatic and for that reason very time consuming. However, recent development in the ultrasound software used at Queen Margaret University have made it possible to use faster ultrasound machines and have improved some of the limitations of the method (such us ultrasound and audio signal alignment) and made data collection and annotation more precise and less laborious. Additionally, a few attempts have been made so far to apply ultrasound to speech therapy and they presented positive outcomes (Shawker and Sonies, 1985; Bernhardt et al., 2003, 2005; Boyce and Schmidlin, 2009). Making ultrasound systems more user-friendly and expanding their applications in speech therapy cause wider interest in the methodology and it can be expected that ultrasound will become more and more present in speech therapy clinics.

Chapter 6

Summary and conclusions

The main aim of this study was to reveal speech characteristics of CAS by applying joint acoustic and articulatory analysis. This was achieved by ultrasound imaging of the tongue which provides articulatory and acoustic information about recorded speech material. In order to reveal unique information about speech in CAS, performance of speakers with CAS was compared to a control group of adults and a control group of typically developing children. The findings did not reveal only information about speech in CAS but also some interesting observations about speech in the two control groups.

As expected, adult speakers showed the shortest syllable durations, the fastest tongue movements over the syllables, and very stable tongue movement patterns in syllable repetitions.

The study confirmed some previously reported differences in the speech of adults and children and provided additional support for the view that in the speech of typically developing children temporal characteristics of syllables stabilise before temporal characteristics of segments. Children aged 6-9 years demonstrated that they already possess mature, adult-like, motor control of the overall tongue movements but still do not differentiate between different parts of the tongue and lack fine motor control over them. They produced similar tongue movements to adults but were more varied in repetitions of the same syllable. Interestingly, the two groups did not differ in the within-group variability of the amount of tongue movement. Such a finding suggests, firstly, that observing the movement of the entire tongue in the described way potentially masks fine differences in individual parts of the tongue while encompassing the movement of the entire tongue and second, that speakers are allowed a relatively large deviation from the "optimal" tongue movement track. The allowed range is large enough that even young children with immature motor-control system produce perceptually correct speech sounds and do not show higher within-group variability than adults.

As a group, speakers with CAS differed from adults on the measures of syllable duration and the rate of tongue movement over syllables, but not from children. They additionally did not differ from the control groups in the amount of tongue movement over the syllable or in the within-group variability on any of these measures. All three speakers adjusted syllable duration to the number of segments in the same way as the AD group and two of them additionally adjusted segment durations in the same way. The latter suggests that some speakers with CAS achieve adult-like control over the investigated temporal features while others perform differently even in the late teenage years. Two of the speakers with CAS also showed the same effect of the number of lingual segments to the amount of tongue movement as the control groups. Ultrasound imaging was particularly valuable for the observation of tongue movement patterns. All three speakers with CAS produced different patterns than the control groups but were very consistent in repetitions.

Speakers with CAS also moved their tongues less in the oral space. It was argued (section 5.3.4) that this characteristic can explain two of the most commonly reported speech characteristics: the presence of a high number of segmental errors, both consonants and vowels, and the presence of groping movements.

Another important aspect of CAS that was addressed in this thesis is inconsistency of production, one of the most commonly reported characteristics of CAS. Surprisingly, the three speakers showed similar or even greater consistency than the control groups in syllable duration, amount and rate of tongue movement over the syllable, and very consistent tongue movement patterns in syllable repetitions. But to answer this question more reliably, a higher number of participants is needed. They should also be evaluated at different times and by producing different kinds of speech material.

The study also raised an interesting idea, that investigating speech characteristics of CAS might benefit from including more than one control group. As a group, participating speakers with CAS performed more similarly to the TDC group but differed from the TDC group in their relationship to the AD group. Even richer information might be obtained by including a group of speakers with similar speech characteristics, such as speakers with phonological disorder.

Finally, this study additionally contributed to the methodological issues related to articulatory analysis of ultrasound data. Firstly, it showed that ultrasound imaging provides valuable information which enables observation of differences in the tongue movement patterns of adults, typically developing children and speakers with CAS. It further confirmed that ultrasound enables measurements of tongue movements from a sequence of speech sounds. Although the main measure used in the study, the amount of tongue movement over a syllable, did not have the ability to clearly differentiate between the speaker groups, it provides a starting point for future development of more complex measures.

The study presented in this thesis contributed to our understanding of CAS by revealing characteristics of tongue movement and temporal features. By including two control groups, it also provided some interesting observations of these features in the speech of adults and typically developing children. Finally, it showed that ultrasound imaging of the tongue can be successfully used to reveal articulatory characteristics of impaired and non-impaired speech. Reported results present a promising direction for future discovery of main and differential characteristics of CAS.

Appendices

Appendix I. Documents related to the recruitment of participants

Queen Margaret University
EDINBURGH

Participant Information Sheet

Tongue movements: Ultrasound study

My name is Tanja Kocjancic and I am a postgraduate student from the School of Social Sciences, Media & Communication at Queen Margaret University in Edinburgh. As part of my research project, I am undertaking a pilot study titled "The influence of syllable complexity on tongue's speed: an ultrasound study".

This study will investigate tongue movements in simple and complex syllables and will serve as a baseline for the future study of tongue movements in developmental apraxia of speech.

I am looking for volunteers to participate in the project. The participants have to be native speakers of English, between 18 and 30 years old, and without any speech impairments.

If you agree to participate in the study, you will be asked to come to the ultrasound recording studio at the Speech Science Research Centre at QMU. You will have to produce six different utterances, each five times. The utterances will be presented on the screen. Before the recording, a specially designed helmet will be placed on your head to stabilize the head and to fix the ultrasound probe. The helmet does not cause any distress, is not uncomfortable and does not intervene with the speech. During the recording the ultrasound probe will be placed under the subjects chin to enable tongue imaging. At the same time the acoustic signal will be recorded via the free standing microphone placed in front of the subject.

The researcher is not aware of any risks associated with the usage of ultrasound. The whole procedure should take no longer than 30 minutes. You will be free to withdraw from the study at any stage and you would not have to give a reason.

All data will be anonymised as much as possible but you may be identifiable from the audio recordings of your voice. Your name will be replaced with a participant number, and it will not be possible for you to be identified in any reporting of the data gathered.

The results of the study may be published in a journal or presented at a conference.

If you would like to contact an independent person, who knows about this project but is not involved in it, you are welcome to contact Dr Jim Scobbie. His contact details are given below.

If you have read and understood this information sheet, any questions you had have been answered, and you would like to be a participant in the study, please now see the consent form.

Queen Margaret University
EDINBURGH

Tongue movements: Ultrasound study

Please read the following statements carefully. Tick all boxes that you agree to and sign below.

1. I confirm that I have read and understood the information sheet for the above study. I have had the opportunity to consider the information, ask questions and have these answered satisfactorily. ☐
2. I understand that my child's participation is voluntary and that my child and/or me are free to withdraw at any time without giving any reason. ☐
3. I have read the consent form and agree that my child takes part in the study. ☐

The next statements relate to the use of your child's audio recordings. Please read and tick all boxes that you agree to.

1. I agree to being **audio-recorded** and that the recordings may be kept for **teaching** purposes. ☐
2. I agree to being **audio-recorded** and that the recordings may be kept for demonstration and **research** purposes. ☐
3. I agree to being **audio-recorded** and that the recordings may be kept for **publication** purposes. ☐

______________________	__________	______________________
Name of Participant	**Date**	**Signature of Participant**
______________________	__________	______________________
Researcher	**Date**	**Signature**

Queen Margaret University
EDINBURGH

Participant Information Sheet

Tongue movements in children: Ultrasound study

My name is Tanja Kocjancic and I am a postgraduate student at the Speech and Hearing Sciences department at Queen Margaret University in Edinburgh. I am working on a project titled: "Ultrasound study of tongue movements in developmental apraxia of speech".

I would like to invite your child to take part in a research study. Before you decide you need to understand why the research is being done and what it would involve for your child and you. Please take time to read the following information carefully and discuss it with your child. If there is anything that is not clear or if you would like more information please feel free to ask us. Take time to decide whether or not you wish to take part.

The aim of this study is to get a better understanding of tongue movements in developmental apraxia of speech (DAS). One of the main features of DAS is impaired articulation causing the speakers to produce wrong speech sounds. Since the tongue is the most important and active speech organ we would like to observe its movements in the speech of teenagers with DAS and in control groups of typically speaking adults and children without DAS. By doing this we will find out if and how do tongue movements in speakers with DAS differ from the other two groups of speakers without DAS. This information could be used in future developing of diagnostic and therapy procedures for speakers with DAS.

As part of the control group I am looking for 6 – 9 years old children to participate in the project. The participants have to be native speakers of English with both parents also being native speakers of English so that we can exclude any influence of the language the child speaks. Because of researching tongue movements in children with typical speech, the participants shouldn't have any known speech, language, hearing or cognitive impairments.

If your child and you agree to participate in the study, you will be asked to come to the ultrasound recording studio at the Speech Science Research Centre at QMU. We will discuss the time of the recording in order to find a time most convenient for you and you will be reimbursed for the travel expenses. You will also receive a £10 WH Smith voucher as a 'thank-you' for the participation.

Tongue movements will be recorded by placing the ultrasound probe under the child's chin. It is very important that the probe does not move during the recording. Because of this, the child will have to wear a special helmet which stabilises the probe under the chin. The helmet does not cause any distress, is not uncomfortable and does not interfere with the speech. As part of the probe fitting some gel will be put on the probe to improve the contact between the probe and the chin. After the fitting of the helmet and the probe, your child will be asked to say different words which will be presented on the screen in front of him/her. During the ultrasound recording we will also record your child's speech.

Ultrasound is a safe and non-invasive method used to scan the inside of the body. The machine we use is the same as the machine used to scan pregnant women to observe the development of the baby. The only difference is that we use it to scan the mouth. The method is used in several speech research labs all over the world and has been used at Queen Margaret University for several years.

Before the recording we will give you and your child a chance to get familiar with the recording studio, the ultrasound machine and the helmet. We will explain the procedure and answer any questions. During the recording you will be allowed to stay with your child in the recording studio.

In addition to the recording we will also ask you for your child's date of birth.

The whole procedure should not take more than 30 minutes since the fitting of the helmet takes up to 10 minutes and the recording takes another 10 minutes. You and your child will be free to withdraw from the study at any stage and you would not have to give a reason.

All data will be anonymised as much as possible but there is a small chance of being recognized from the audio recordings of the speech. It is not however possible to be identified from the ultrasound recordings since they show only the outline of the tongue. Your child's name will be replaced with a participant number and it will not be possible for any participant to be identified in any reporting of the data gathered. All the recordings will be stored on a password protected computer.

The data and results of the study will be published in my PhD thesis, published in a journal or presented at a conference. If you agree, the audio recordings will be used for presentations, publications and teaching purpose as well.

I would additionally like to inform you that this is not an assessment of your child's speech. We will only look at tongue movement and will not give any opinion about speech development or potential speech impairments.

If you have any more questions please contact me or an independent person, who knows about this project but is not involved in it, Dr Jim Scobbie. Our contact details are given below.

If you have read and understood this information sheet, any questions you had have been answered, and you and your child agree to participate in the study, please now see the consent form.

Queen Margaret University
EDINBURGH

Tongue movements in children: Ultrasound study

CONSENT FORM

Please read the following statements carefully. Tick all boxes that you agree to and sign below.

1. I confirm that I have read and understood the information sheet for the above study. I have had the opportunity to consider the information, ask questions and have these answered satisfactorily. ☐
2. I understand that my child's participation is voluntary and that my child and/or me are free to withdraw at any time without giving any reason. ☐
3. I have read the consent form and agree that my child takes part in the study. ☐

The next statements relate to the use of your child's audio recordings. Please read and tick all boxes that you agree to.

1. I agree that my child is **audio-recorded** and that the recordings may be kept for **teaching** purposes. ☐
2. I agree that my child is **audio-recorded** and that the recordings may be kept for demonstration and **research** purposes. ☐
3. I agree that my child is **audio-recorded** and that the recordings may be kept for **publication** purposes. ☐

____
Name of Participant

______________________	____________	______________________
Name of Participant's Parent	**Date**	**Signature of Participant's Parent or Guardian**
______________________	____________	______________________
Researcher	**Date**	**Signature**

Queen Margaret University
EDINBURGH

Participant Information Sheet – Adults

Project Title:
Ultrasound investigation of tongue movements in developmental apraxia of speech

Principal Investigator: Tanja Kocjancic

We would like to invite you to take part in a research study. Before you decide you need to understand why the research is being done and what it would involve for you. Please take time to read the following information carefully and discuss it with others if you wish. If there is anything that is not clear or if you would like more information please feel free to ask us. Take time to decide whether or not you wish to take part.

What is the aim of the study?

The aim of this study is to get a better understanding of tongue movements in developmental apraxia of speech (DAS). This experiment is a part of the principal investigator's research, in order to obtain a PhD degree at Queen Margaret University, Edinburgh.

Impaired articulation is one of the main features of DAS. Since the tongue is the most important and active articulator in speech we would like to observe its movements in DAS and compare them to tongue movements of young adults and children without DAS. Data has already been collected from adult control speakers in an independent study and data from children will be recorded in the future.

In a separate part of the study a video recording of your spontaneous speech will be played to a group of listeners to be rated on intelligibility and be assessed for DAS features.

Who will benefit?

Speakers with DAS may benefit from the study. The results of the study could be used in designing diagnostic procedures using ultrasound. Furthermore, they may reveal some features of tongue movements that could be used in designing therapy procedures.

Why have I been invited?

You have been invited following a recommendation by Professor Anne O'Hare at Community Child Health, Royal Hospital for Sick Children, because you satisfy the inclusion criteria for taking part in this research. We would like to record ten speakers with DAS.

Do I have to participate?

No. It is up to you to decide. We will describe the study to you and answer all your questions. If you decide to participate we will ask you to sign a consent form to show you have agreed to take part. You are free to withdraw at any time, without giving a reason. Deciding to participate or not to participate, or to withdraw from the study, will not affect the standard of care you receive.

What will happen to me if I take part?

To get all the necessary data three recording techniques will be used: ultrasound, audio, and video.

Ultrasound and audio recording

Ultrasound is a safe and non-invasive method used to scan the inside of the body. The machine we use is the same as the machine used to scan pregnant women to observe the development of the baby. The only difference is that we use it to scan the mouth. The method is used in several speech research labs all over the world and has been used at Queen Margaret University for several years.

Tongue movements will be recorded by placing the ultrasound probe under your chin. It is very important that the probe does not move during the recording. Because of that, you will have to wear a special helmet (Figure 1) which stabilizes the probe under your chin. Additionally, some gel will be put on the probe to improve the contact between the probe and your chin. After the fitting of the helmet and the probe, you will be asked to say different words which will be presented on the screen in front of you. During the ultrasound recording we will also record your speech.

The fitting of the helmet takes up to 10 minutes and the recording takes another 10 minutes.

Figure 1: Ultrasound helmet

Video recording

Because DAS is a motor speech disorder we would like to observe your speech motor skills as well. We will ask you to perform several tasks involving the movement of your mouth and tongue. In addition, we will also record a short sample (2 minutes) of your spontaneous speech.

The video recording will last 10 minutes in total.

What will I have to do?

We will ask you to come once to the ultrasound studio at Queen Margaret University, Edinburgh. We will discuss the time of the recording in order to find a time most convenient for you and you will be reimbursed for the travel expenses. You will also receive a £10 WH Smith voucher as a 'thank-you' for the participation. The visit will not last more than one hour.

During the visit you will:

1. Perform different speech motor tasks and speak with the researcher to record some of your spontaneous speech. This part will be video recorded.
2. Say the words presented on the screen. This will be audio and ultrasound recorded as described above.

What will we do with the data?

All the data will be anonymised and analysed only in such form. Your name will be replaced by a code such as speaker S1, S2...

In order to assess intelligibility and the characteristics of DAS, short video samples of your spontaneous speech will be played to speech and language therapy students at Queen Margaret University.

Additional data, such as date of birth, year of diagnosing DAS and the duration of speech therapy will be gathered by Dr Anne O'Hare and kept in accordance with data protection requirements.

The results of the study will be published in the principal investigator's PhD thesis, in appropriate scientific journals and presented at relevant scientific conferences. Your name will not be published or presented in any case and you can not be identified from the ultrasound recordings.

If you agree we would like to ask for the permission to use short clips of your audio and video recordings as a demonstration in presentations and publications, and for teaching purposes. You will have a chance to decide on each of these options on the consent form.

You can still participate in this study, even if you do not want us to use your recordings for any purposes other than those directly connected to the project.

The report of the study will be available upon request from the principal investigator.

What are the possible disadvantages and risks of taking part?

There are no health or safety risks associated with ultrasound. There is a small possibility of a discomfort when wearing the helmet. We will address any discomfort immediately and readjust the helmet. Because the helmet is fixed to the head there is a chance of some slight reddening of the skin for a short while after the removal of the helmet.

You can not be recognized from the ultrasound images of your tongue. However, you can be recognized from video recordings or from the audio recordings of your speech. Your name will not be present in any of the recordings and the recordings will only be used for the research and teaching purpose.

Consent form

Please read the consent form carefully. The consent form allows you to express your decision about participating in the study and about the use of the recordings. It also informs you that you can withdraw from the study at any time without giving any reason.

It is a good practice to inform your GP about your participation in any health related study. If you agree we will send your GP a letter explaining your participation.

Before the recordings you will sign two consent forms. One will stay with you and one will be kept in our records.

Further information

If you have any further information please contact the members of the research team or the independent advisor. The contact details are below.

Queen Margaret University
EDINBURGH

Project Title:
Ultrasound investigation of tongue movements in developmental apraxia of speech

Principal Investigator: Tanja Kocjancic

CONSENT FORM - Adults

Please read the following statements carefully. Tick all boxes that apply and sign at the end of the form.

	YES	NO
1. I confirm that I have read and understood the information sheet for the above study. I have had the opportunity to consider the information, ask questions and have these answered satisfactorily.	☐	☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason.	☐	☐
3. I agree that my General Practitioner (GP) will be informed of my participation in the study.	☐	☐
4. I have read the consent form and agree to take part in the study.	☐	☐

Please turn and continue on the next page...

The next statements relate to the use of your audio and video recordings. Please read and tick all the boxes that apply.

	YES	NO
1. I agree to being **audio-recorded** and that the recordings may be kept for **research** purposes.	☐	☐
2. I agree to being **audio-recorded** and that the recordings may be kept for **scientific presentation** purposes.	☐	☐
3. I agree to being **audio-recorded** and that the recordings may be kept for **teaching** purposes.	☐	☐
4. I agree to being **audio-recorded** and that the recordings may be kept for **publication** purposes.	☐	☐
5. I agree to being **video-recorded** and that the recordings may be kept for **research** purposes.	☐	☐
6. I agree to being **video-recorded** and that the recordings may be kept for **scientific presentation** purposes.	☐	☐
7. I agree to being **video-recorded** and that the recordings may be kept for **teaching** purposes.	☐	☐

________________________	__________	________________________
Name of Participant	**Date**	**Signature of Participant**
________________________	__________	________________________
Researcher	**Date**	**Signature**

Thank you for your help.

Queen Margaret University
EDINBURGH

Participant Information Sheet – Young person

Project Title:
Ultrasound investigation of tongue movements in developmental apraxia of speech

Principal Investigator: Tanja Kocjancic

We would like to invite you to take part in a research study. Before you decide you need to understand why the research is being done and what it would involve for you. Please take time to read the following information carefully and discuss it with your family and friends if you wish. If there is anything that is not clear or if you would like more information please feel free to ask us. Take time to decide whether or not you wish to take part.

What is the aim of the study?

The aim of this study is to get a better understanding of how the tongue moves in the speech of speakers with developmental apraxia of speech (DAS). This experiment is a part of the principal investigator's research, in order to obtain a PhD degree at Queen Margaret University, Edinburgh.

Impaired articulation is one of the main features of DAS. Since the tongue is the most important and active articulator in speech we would like to observe its movements in DAS and compare them to the tongue movements of young adults and children without DAS. Data has already been collected from adult control speakers in an independent study and data from children will be recorded in the future.

In a separate part of the study a video recording of your spontaneous speech will be played to a group of listeners to be rated on intelligibility and be assessed for DAS features.

Who will benefit?

Speakers with DAS may benefit from the study. The results of the study could be used in designing diagnostic procedures using ultrasound. Furthermore, they may reveal some features of tongue movements that could be used in designing therapy procedures.

Why have I been invited?

You have been invited to our study because you have DAS and have been recommended by Professor Anne O'Hare at Community Child Health, Royal Hospital for Sick Children. We would like to record ten speakers with DAS.

Do I have to participate?

No. It is up to you. If you do, the researcher will ask you to sign a form giving your consent or assent. You will be given a copy of this information sheet and your signed form to keep. You are free to stop taking part at any time during the research without giving a reason. If you decide to stop, this will not affect the care you receive.

What will happen to me if I take part?

To get all the necessary data three recording techniques will be used: ultrasound, audio, and video.

Ultrasound and audio recording

Ultrasound is a safe and non-invasive method used to scan the inside of the body. The machine we use is the same as the machine used to scan pregnant women to observe the development of the baby. The only difference is that we use it to scan the mouth. The method is used in several speech research labs all over the world and has been used at Queen Margaret University for several years.

Tongue movements will be recorded by placing the ultrasound probe under your chin. It is very important that the probe does not move during the recording. Because of that, you will have to wear a special helmet which stabilizes the probe under your chin. Additionally, some gel will be put on the probe to improve the contact between the probe and your chin. After the fitting of the helmet and the probe, you will be asked to say different words which will be presented on the screen in front of you. During the ultrasound recording we will also record your speech.

The fitting of the helmet takes up to 10 minutes and the recording takes another 10 minutes.

Figure 1: Ultrasound helmet

Video recording

Because DAS is a motor speech disorder we would like to observe your speech motor skills as well. We will ask you to perform several tasks involving the movement of your mouth and tongue. In addition, we will also record a short sample (2 minutes) of your spontaneous speech.

The video recording will last 10 minutes in total.

What will I have to do?

We will ask you to come once to the ultrasound studio at Queen Margaret University, Edinburgh. We will discuss the time of the recording in order to find a time most convenient for you and you will be reimbursed for the travel expenses. You will also receive a £10 WH Smith voucher as a 'thank-you' for the participation. The visit will not last more than one hour.

During the visit you will:

1. Perform different speech motor tasks and speak with the researcher to record some of your spontaneous speech. This part will be video recorded.
2. Say the words presented on the screen. This will be audio and ultrasound recorded as described above.

What will we do with the data?

All the data will be anonymised and analysed only in such form. Your name will be replaced by a code such as speaker S1, S2...

In order to assess intelligibility and the characteristics of DAS, short video samples of your spontaneous speech will be played to speech and language therapy students at Queen Margaret University.

Additional data, such as date of birth, year of diagnosing DAS and the duration of speech therapy will be gathered by Dr Anne O'Hare and kept in accordance with data protection requirements.

The results of the study will be published in the principal investigator's PhD thesis, in appropriate scientific journals and presented at relevant scientific conferences. Your name will not be published or presented in any case and you can not be identified from the ultrasound recordings.

If you agree we would like to ask for the permission to use short clips of your audio and video recordings as a demonstration in presentations and publications, and for teaching purposes. You will have a chance to decide on each of these options on the consent form.

You can still participate in this study, even if you do not want us to use your recordings for any purposes other than those directly connected to the project.

If you would like to know about the results of this study you can ask the researchers to send you a short report.

What are the possible disadvantages and risks of taking part?

There are no health or safety risks associated with ultrasound. There is a small possibility of a discomfort when wearing the helmet. We will address any discomfort immediately and readjust the helmet. Because the helmet is fixed to the head there is a chance of some slight reddening of the skin for a short while after the removal of the helmet.

You can not be recognized from the ultrasound images of your tongue. However, you can be recognized from video recordings or from the audio recordings of your speech. Your name will not be present in any of the recordings and the recordings will only be used for the research and teaching purpose.

Consent and Assent forms

Please read the consent form carefully with your parents. The consent form allows your parents, who are your legal representatives, to tell us if you wish to participate in the study and how would you like us to use your recordings. It also informs you that you can withdraw from the study at any time without giving any reason.

Additionally we are asking you to carefully read the assent form as well. Assent form gives us additional information that you have understood what the project is about an that you would like to take part.

It is a good practice to inform your GP about your participation in any health related study. If you and your parents agree we will send your GP a letter explaining your participation.

Before the recordings one of your parents will sign two consent forms and you will sign two assent forms. One of each will stay with you and your parents and one will be kept in our records.

Further information

If you have any further information please contact the members of the research team or the independent advisor. The contact details are below.

Queen Margaret University
EDINBURGH

Project Title:
Ultrasound investigation of tongue movements in developmental apraxia of speech

Principal Investigator: Tanja Kocjancic

CONSENT FORM – Young person

Please read the following statements carefully. Tick all boxes that apply and sign at the end of the form.

	YES	NO
1. I confirm that I have read and understood the information sheet for the above study. I have had the opportunity to consider the information, ask questions and have these answered satisfactorily.	☐	☐
2. I understand that my child's participation is voluntary and that I am free to withdraw at any time without giving any reason.	☐	☐
3. I agree that my child's General Practitioner (GP) will be informed of my child's participation in the study.	☐	☐
4. I have read the consent form and agree that my child takes part in the study.	☐	☐

Please turn and continue on the next page...

The next statements relate to the use of your child's audio and video recordings. Please read and tick all the boxes that apply.

	YES	NO
1. I agree that my child is **audio-recorded** and that the recordings may be kept for **research** purposes.	☐	☐
2. I agree that my child is **audio-recorded** and that the recordings may be kept for **scientific presentation** purposes.	☐	☐
3. I agree that my child is **audio-recorded** and that the recordings may be kept for **teaching** purposes.	☐	☐
4. I agree that my child is **audio-recorded** and that the recordings may be kept for **publication** purposes.	☐	☐
5. I agree that my child is **video-recorded** and that the recordings may be kept for **research** purposes.	☐	☐
6. I agree that my child is **video-recorded** and that the recordings may be kept for **scientific presentation** purposes.	☐	☐
7. I agree that my child is **video-recorded** and that the recordings may be kept for **teaching** purposes.	☐	☐

____
Name of Participant

______________________________ ____________ ______________________________
Name of Participant's Parent **Date** **Signature of Participant's Parent**

______________________________ ____________ ______________________________
Researcher **Date** **Signature**

Thank you for your help.

Queen Margaret University
EDINBURGH

Project Title:
Ultrasound investigation of tongue movements in developmental apraxia of speech

Principal Investigator: Tanja Kocjancic

ASSENT FORM – Young person

Please read the following statements and circle all you agree with.

Have you read about this project?	Yes / No
Has somebody else explained this project to you?	Yes / No
Do you understand what this project is about?	Yes / No
Have you asked all the questions you want?	Yes / No
Have you had your questions answered in a way you understand?	Yes / No
Do you understand it's OK to stop taking part at any time?	Yes / No
Are you happy to take part?	Yes / No

If any answers are 'no' or you don't want to take part, don't sign your name!

If you do want to take part, you can write your name below

Your name ______________________

Date ______________________

The researcher who explained this project to you needs to sign too

Print name ______________________

Sign ______________________

Date ______________________

Thank you for your help.

Appendix II. Results of statistical tests

Measure		dur		aom		rom	
Syllable	N	W	*p*	W	*p*	W	*p*
pay	50	0.973	0.293	0.977	0.419	0.945	.021
say	50	0.968	0.193	0.929	0.005	0.983	0.672
lay	50	0.953	0.044	0.934	0.008	0.835	0.000
play	50	0.941	0.014	0.932	0.006	0.845	0.000
slay	50	0.972	0.278	0.978	0.466	0.884	0.000
splay	50	0.975	0.368	0.971	0.253	0.844	0.000

Table 1: Shapiro-Wilk test of normality evaluating distribution of syllable duration (dur), amount of tongue movement (aom) and rate of tongue movement (rom) by syllable type in group AD.

Measure		dur		aom		rom	
Syllable	N	W	*p*	W	*p*	W	*p*
pay	50	0.926	0.004	0.961	0.098	0.914	0.001
say	50	0.985	0.769	0.984	0.712	0.980	0.539
lay	50	0.952	0.041	0.978	0.463	0.934	0.008
play	50	0.967	0.169	0.982	0.649	0.947	0.027
slay	50	0.974	0.329	0.984	0.720	0.940	0.013
splay	50	0.961	0.096	0.977	0.451	0.926	0.004

Table 2: Shapiro-Wilk test of normality evaluating distribution of syllable duration (dur), amount of tongue movement (aom) and rate of tongue movement (rom) by syllable type in group TDC.

Measure		dur		aom		rom	
Syllable	N	W	*p*	W	*p*	W	*p*
pay	15	0.859	0.023	0.856	0.021	0.970	0.857
say	15	0.930	0.275	0.945	0.448	0.946	0.469
lay	15	0.892	0.071	0.950	0.525	0.938	0.362
play	15	0.881	0.048	0.902	0.104	0.934	0.312
slay	15	0.823	0.007	0.911	0.139	0.875	0.040
splay	15	0.966	0.802	0.879	0.045	0.932	0.288

Table 3: Shapiro-Wilk test of normality evaluating distribution of syllable duration (dur), amount of tongue movement (aom) and rate of tongue movement (rom) by syllable type in group CAS.

	Speakers	AD		TDC		CAS	
Segment	Syllable	Median	IQR	Median	IQR	Median	IQR
/p/	pay	157	29	184	76	192	26
	play	174	41	181	72	193	24
	splay	106	16	123	32	123	26
/s/	say	170	54	217	108	206	23
	slay	182	40	225	95	217	32
	splay	121	32	152	76	147	22
/l/	lay	82	39	111	73	107	23
	play	26	14	42	20	27	20
	slay	47	27	62	53	62	21
	splay	43	23	79	57	68	19

Table 4: Median and IQR values of /p/, /s/, and /l/ duration (ms) by syllable type and speaker group.

Syllable pair	Measure	N	Z	*p*
pay-play	dur	10	-2.803	0.0025
	aom	10	-2.803	0.0025
	rom	10	-2.191	0.0140
say-slay	dur	10	-2.803	0.0025
	aom	10	-2.803	0.0025
	rom	10	-2.395	0.0085
lay-play	dur	10	-2.803	0.0025
	aom	10	-2.701	0.0035
	rom	10	-2.599	0.0045
lay-slay	dur	10	-2.803	0.0025
	aom	10	-2.191	0.0140
	rom	10	-2.701	0.0035
play-splay	dur	10	-2.803	0.0000
	aom	10	-1.784	0.0370
	rom	10	-2.293	0.0110
slay-splay	dur	10	-2.191	0.0140
	aom	10	-2.191	0.0140
	rom	10	-1.376	0.0845

Table 5: Wilcoxon Signed-Rank test evaluating differences in syllable duration (dur), amount of tongue movement (aom) and rate of tongue movement (rom), between syllable pairs in group AD.

Syllable pair	Measure	N	Z	*p*
pay-play	dur	10	-2.293	0.0110
	aom	10	-2.599	0.0045
	rom	10	-2.191	0.0140
say-slay	dur	10	-2.599	0.0045
	aom	10	-2.803	0.0025
	rom	10	-2.293	0.0085
lay-play	dur	10	-2.803	0.0025
	aom	10	-2.803	0.0025
	rom	10	-0.153	0.0045
lay-slay	dur	10	-2.803	0.0025
	aom	10	-2.803	0.0025
	rom	10	-2.803	0.0025
play-splay	dur	10	-2.803	0.0025
	aom	10	-1.274	0.1015
	rom	10	-2.803	0.0100
slay-splay	dur	10	-2.293	0.0110
	aom	10	-1.070	0.1425
	rom	10	-0.561	0.0845

Table 6: Wilcoxon Signed-Rank test evaluating differences in syllable duration (dur), amount of tongue movement (aom) and rate of tongue movement (rom), between syllable pairs in group TDC.

Syllable pair	Measure	N	Z	*p*
pay-play	dur	3	-1.604	0.1090
	aom	3	-1.604	0.1090
	rom	3	-1.604	0.1090
say-slay	dur	3	-1.604	0.1090
	aom	3	-1.604	0.1090
	rom	3	-0.535	0.5930
lay-play	dur	3	-1.604	0.1090
	aom	3	-1.604	0.1090
	rom	3	-1.604	0.1090
lay-slay	dur	3	-1.604	0.1090
	aom	3	-1.604	0.1090
	rom	3	-1.604	0.1090
play-splay	dur	3	-1.604	0.1090
	aom	3	-1.069	0.2850
	rom	3	-1.604	0.1090
slay-splay	dur	3	-1.604	0.1090
	aom	3	-1.604	0.1090
	rom	3	-0.535	0.5930

Table 7: Wilcoxon Signed-Rank test evaluating differences in syllable duration (dur), amount of tongue movement (aom) and rate of tongue movement (rom), between syllable pairs in group CAS.

Segment	Syllable pair	N	Z	*p*
/p/	pay-play	10	-2.292	0.0110
	pay-splay	10	-2.803	0.0025
/s/	say-slay	10	-2.497	0.0065
	say-splay	10	-2.803	0.0025
/l/	lay-play	10	-2.803	0.0025
	lay-slay	10	-2.701	0.0035
	lay-splay	10	-2.803	0.0025

Table 8: Wilcoxon Signed-Rank test evaluating differences in segment durations between syllable pairs in group AD.

Segment	Syllable pair	N	Z	*p*
/p/	pay-play	10	-2.299	0.0695
	pay-splay	10	-2.803	0.0025
/s/	say-slay	10	-2.497	0.0140
	say-splay	10	-2.803	0.0025
/l/	lay-play	10	-2.803	0.0025
	lay-slay	10	-2.701	0.0035
	lay-splay	10	-2.803	0.0065

Table 9: Wilcoxon Signed-Rank test evaluating differences in segment durations between syllable pairs in group TDC.

Segment	Syllable pair	N	Z	*p*
/p/	pay-play	3	-1.069	0.2850
	pay-splay	3	-1.604	0.1090
/s/	say-slay	3	-1.604	0.1090
	say-splay	3	-1.604	0.1090
/l/	lay-play	3	-1.604	0.1090
	lay-slay	3	-1.604	0.1090
	lay-splay	3	-1.604	0.1090

Table 10: Wilcoxon Signed-Rank test evaluating differences in segment durations between syllable pairs in group CAS.

Measure		dur			aom			rom		
Syllable	N	U	Z	*p*	U	Z	*p*	U	Z	*p*
pay	10:10	22	-2.117	0.0170	48	-0.151	0.4400	10	-3.024	0.0010
say	10:10	19	-2.343	0.0065	37	-0.983	0.1630	10	-0.024	0.0010
lay	10:10	22	-2.117	0.0170	23	-2.041	0.0205	13	-2.797	0.0025
play	10:10	24	-1.965	0.0245	39	-0.832	0.2030	20	-2.268	0.0115
slay	10:10	17	-2.495	0.0065	38	-0.907	0.1820	18	-2.1419	0.0080
splay	10:10	18	-2.419	0.0080	39	-0.832	0.2030	17	-2.495	0.0065

Table 11: Mann-Whitney test evaluating differences in syllable duration (dur), amount of tongue movement (aom) and rate of tongue movement (rom), between syllable types in groups AD and TDC.

Measure		dur			aom			rom		
Syllable	N	U	Z	*p*	U	Z	*p*	U	Z	*p*
pay	10:10	16	-2.570	0.0050	41	-0.680	0.2480	27	-1.739	0.0041
say	10:10	32	-1.361	0.0870	31	-1.437	0.0755	47	-0.227	0.4105
lay	10:10	12	-2.873	0.0020	42	-0.605	0.2725	48	-0.151	0.4400
play	10:10	11	-2.948	0.0015	39	-0.832	0.2030	48	-0.151	0.4400
slay	10:10	34	-1.209	0.1130	34	-1.209	0.1130	46	-0.302	0.3810
splay	10:10	3	-3.553	0.0000	46	-0.302	0.3810	47	-0.227	0.4105

Table 12: Mann-Whitney test evaluating differences in the variability (IQR) of syllable duration (dur), amount of tongue movement (aom) and rate of tongue movement (rom), between syllable types in groups AD and TDC.

Measure		dur			aom			rom		
Syllable	N	U	Z	*p*	U	Z	*p*	U	Z	*p*
pay	10:3	7	-1.352	0.1760	14	-0.169	0.8660	0	-2.535	0.0110
say	10:3	9	-1.352	0.1760	14	-0.169	0.8660	6	-0.521	0.1280
lay	10:3	9	-1.014	0.3100	9	-1.014	0.3100	3	-2.028	0.0430
play	10:3	8	-1.183	0.2370	12	-0.507	0.6120	1	-2.366	0.0180
slay	10:3	7	-1.352	0.1760	10	-0.845	0.3980	4	-1.859	0.0630
splay	10:3	4	-1.859	0.0630	11	-0.676	0.4990	3	-2.028	0.0430

Table 13: Mann-Whitney test evaluating differences in syllable duration (dur), amount of tongue movement (aom) and rate of tongue movement (rom), between syllable types in groups AD and CAS.

Measure		dur			aom			rom		
Syllable	N	U	Z	*p*	U	Z	*p*	U	Z	*p*
pay	10:3	7	-1.352	0.0880	11	-0.676	0.2495	4	-1.859	0.0315
say	10:3	9	-1.014	0.1550	11	-0.677	0.2490	13	-0.338	0.3675
lay	10:3	14	-0.169	0.4330	6	-1.521	0.0640	11	-0.676	0.2495
play	10:3	10	-0.845	0.1990	15	0.000	0.5000	14	-0.169	0.4330
slay	10:3	14	-0.169	0.4330	8	-1.183	0.1185	14	-0.169	0.4330
splay	10:3	2	-2.197	0.0140	5	-1.690	0.0455	7	-1.352	0.0880

Table 14: Mann-Whitney test evaluating differences in the variability (IQR) of syllable duration (dur), amount of tongue movement (aom) and rate of tongue movement (rom), between syllable types in groups AD and CAS.

Measure		dur			aom			rom		
Syllable	N	U	Z	*p*	U	Z	*p*	U	Z	*p*
pay	10:3	14	-0.169	0.8660	15	0.000	1.0000	15	0.000	1.0000
say	10:3	10	-0.845	0.3980	10	-0.845	0.3980	8	-1.183	0.2370
lay	10:3	11	-0.676	0.4990	14	-0.169	0.8660	13	-0.338	0.7350
play	10:3	12	-0.507	0.6120	14	-0.169	0.8660	12	-0.507	0.6120
slay	10:3	9	-1.014	0.3100	13	-0.338	0.7350	12	-0.507	0.6120
splay	10:3	11	-0.676	0.4990	14	-0.169	0.8660	12	-0.507	0.6120

Table 15: Mann-Whitney test evaluating differences in syllable duration (dur), amount of tongue movement (aom) and rate of tongue movement (rom), between syllable types in groups TDC and CAS.

Measure		dur			aom			rom		
Syllable	N	U	Z	*p*	U	Z	*p*	U	Z	*p*
pay	10:3	9	-1.014	0.1550	10	-0.845	0.1990	9	-1.014	0.1550
say	10:3	14	-0.169	0.4330	15	0.000	0.5000	13	-0.338	0.3675
lay	10:3	3	-2.028	0.0215	5	-1.690	0.0455	8	-1.183	0.1185
play	10:3	4	-1.859	0.0315	14	-0.169	0.4330	14	-0.169	0.4330
slay	10:3	10	-0.845	0.1990	3	-2.028	0.0215	14	-0.169	0.4330
splay	10:3	9	-1.014	0.1550	1	-2.366	0.0090	5	-1.690	0.0455

Table 16: Mann-Whitney test evaluating differences in the variability (IQR) of syllable duration (dur), amount of tongue movement (aom) and rate of tongue movement (rom), between syllable types in groups TDC and CAS.

Fixed effect	Speakers	Estimate	MCMC mean	HPD 95% lower	HPD 95% upper	pMCMC
Number of onset segments	AD	1.5870	1.5820	0.8648	2.3190	0.0014
	TDC	2.0240	2.0160	0.8115	3.1430	0.0032
	CAS	1.8370	1.8460	1.0650	2.5840	0.0001

Table 17: Results of GLMM modelling of the effect of the number of syllable onset segments on syllable durations in each of the speaker groups.

Speakers	Syllable	Estimate	MCMC mean	HPD 95% lower	HPD 95% upper	pMCMC
AD vs. TDC	pay	2.3470	2.3400	1.1870	3.4900	0.0030
	say	2.9320	2.9360	1.8208	4.0810	0.0001
	lay	2.2840	2.2890	1.1620	3.3610	0.0008
	play	2.3780	2.3860	1.2628	3.4400	0.0002
	slay	3.299	3.2980	2.1604	4.5400	0.0002
	splay	2.4540	3.4530	2.1400	4.7220	0.0001
AD vs. CAS	pay	1.8440	1.8400	0.1950	3.5960	0.0360
	say	1.9170	1.9130	0.2671	3.5760	0.0276
	lay	1.3290	1.3430	-0.3120	2.9530	0.1060
	play	1.5240	1.5230	-0.1208	3.0630	0.0636
	slay	1.9580	1.9440	0.1897	3.7500	0.0318
	splay	2.2980	2.2950	0.2694	4.1620	0.0212
TDC vs. CAS	pay	-0.5031	-0.4900	-2.1700	1.1980	0.5620
	say	-1.0150	-1.0200	-2.5710	0.7440	0.2264
	lay	-0.9555	-0.9497	-2.5020	0.7797	0.2570
	play	-0.8549	-0.8520	-2.5320	0.6895	0.2938
	slay	-1.3410	-1.3360	-3.0810	0.4306	0.1332
	splay	-1.1570	-1.1490	-3.0530	0.8117	0.2482

Table 18: Results of GLMM modelling of the effect of speaker group on syllable durations for each of the syllable types.

Speakers	Syllable	Estimate	MCMC mean	HPD 95% lower	HPD 95% upper	pMCMC
AD vs. TDC	pay	2.5490	2.5590	0.8236	4.4070	0.0082
	say	2.0600	2.0606	-0.1058	4.2310	0.0624
	lay	2.6730	2.6695	0.8984	4.3910	0.0052
	play	2.2310	2.2293	0.9105	3.6050	0.0016
	slay	1.5945	1.6016	-0.3903	3.5490	0.1096
	splay	6.3770	6.3570	4.1394	8.6140	0.0001
AD vs. CAS	pay	1.1300	1.1640	-1.4254	3.8650	0.5590
	say	0.9134	0.9051	-2.2007	4.1700	0.5590
	lay	0.4495	0.4386	-2.2396	3.0060	0.7324
	play	0.4300	0.4294	-1.4989	2.4370	0.6576
	slay	-0.3389	-0.3466	-3.4204	2.4510	0.8220
	splay	3.5630	3.5680	0.2625	6.77080	0.0308
TDC vs. CAS	pay	-1.4190	-1.3940	-4.0630	1.3172	0.2874
	say	-1.1470	-1.1500	-4.3431	1.9833	0.4662
	lay	-2.2230	-2.2120	-4.8260	0.3892	0.0904
	play	-1.8010	-1.8080	-3.8460	0.0769	0.0686
	slay	-1.9330	-1.9640	-5.0530	0.9306	0.1850
	splay	-2.8140	-2.8210	-6.0200	0.4294	0.0858

Table 19: Results of GLMM modelling of the effect of speaker group on the variability (IQR) of syllable durations.

Speakers	Segment	Estimate	MCMC mean	HPD 95% lower	HPD 95% upper	pMCMC
AD	/p/	-1.0400	-1.0380	-2.8650	0.9688	0.2620
	/s/	-0.8290	-0.8254	-2.8060	1.0150	0.3506
	/l/	-2.9950	-2.9990	-5.0970	-0.8981	0.0132
TDC	/p/	-1.3010	-1.3120	-4.4530	1.9520	0.0341
	/s/	-1.0390	-1.0450	-4.3710	2.2040	0.4680
	/l/	-3.0080	-3.0030	-5.7700	-0.1325	0.0406
CAS	/p/	-1.2030	-1.1960	-3.3310	0.9298	0.2522
	/s/	-1.0150	-1.0180	-2.8000	0.6858	0.2308
	/l/	-3.2570	-3.2600	-5.7740	-0.7729	0.0182

Table 20: Results of GLMM modelling of the effect of the type of syllable onset (single segment or cluster) on segment durations in each of the speaker groups.

Fixed effect	Speakers	Estimate	MCMC mean	HPD 95% lower	HPD 95% upper	pMCMC
Number of lingual onset segments	AD	0.2428	0.2440	0.0172	0.4613	0.0332
	TDC	0.2366	0.2371	0.0061	0.4643	0.0456
	CAS	0.1844	0.1844	0.0376	0.3238	0.1800

Table 21: Results of GLMM modelling of the effect of the number of lingual syllable onset segments on the amount of tongue movement over the syllable in each of the speaker groups.

Speakers	Syllable	Estimate	MCMC mean	HPD 95% lower	HPD 95% upper	pMCMC
AD vs. TDC	pay	-0.0525	-0.5250	-0.2218	0.1163	0.5354
	say	-0.1739	-0.1748	-0.3407	-0.0051	0.0000
	lay	-0.2888	-0.2886	-0.4669	-0.1225	0.0000
	play	-0.1365	-0.1360	-0.3050	0.0276	0.1034
	slay	-0.0902	-0.0905	-0.2739	0.0972	0.3286
	splay	-0.1242	-0.1237	-0.3142	0.0706	0.1928
AD vs. CAS	pay	-0.0976	-0.0981	-0.3449	0.1438	0.4344
	say	0.0374	0.0377	-0.2107	0.2782	0.7636
	lay	-0.2868	-0.2835	-0.5429	-0.0291	0.0000
	play	-0.2363	-0.2360	-0.4862	-0.0048	0.0544
	slay	-0.2073	-0.2052	-0.4795	0.0573	0.1324
	splay	-0.2250	-0.2553	-0.5178	0.0675	0.1206
TDC vs. CAS	pay	-0.0452	-0.0449	-0.2936	0.2008	0.7196
	say	0.2110	0.2117	-0.0319	0.4658	0.0898
	lay	0.0020	0.0030	-0.2475	0.2627	0.9880
	play	-0.0998	-0.1002	-0.3480	0.1341	0.4134
	slay	-0.1172	-0.1146	-0.3996	0.1623	0.3842
	splay	-0.1008	-0.0986	-0.3742	0.1883	0.4728

Table 22: Results of GLMM modelling of the effect of the speaker group on the amount of tongue movement over syllable for each of the syllable types.

Speakers	Syllable	Estimate	MCMC mean	HPD 95% lower	HPD 95% upper	pMCMC
AD vs. TDC	pay	0.0549	0.0576	-0.1657	0.2757	0.5882
	say	0.2202	0.2209	-0.0656	0.5091	0.1258
	lay	0.0993	0.1000	-0.1364	0.3319	0.3956
	play	0.0359	0.0351	-0.2131	0.2687	0.7674
	slay	0.1105	0.1118	-0.1240	0.3322	0.3236
	splay	0.0734	0.0751	-0.2105	0.3872	0.6054
AD vs. CAS	pay	-0.0834	-0.0830	-0.4033	0.2511	0.5884
	say	0.2381	0.2383	-0.2114	0.6492	0.2586
	lay	-0.2597	-0.2608	-0.6183	0.0964	0.1430
	play	-0.0097	-0.0080	-0.3581	0.3506	0.9684
	slay	-0.1629	-0.1627	-0.5060	0.1711	0.3386
	splay	-0.3685	-0.3680	-0.8086	0.0717	0.0974
TDC vs. CAS	pay	-0.1384	-0.1358	-0.4608	0.1971	0.4108
	say	0.0179	0.0138	-0.4388	0.4312	0.9424
	lay	-0.3591	-0.3599	-0.7174	-0.0063	0.0476
	play	-0.0456	-0.0494	-0.4147	0.3057	0.7872
	slay	-0.2735	-0.2709	-0.5998	0.0614	0.1084
	splay	-0.4419	-0.4430	-0.8870	-0.0072	0.0486

Table 23: Results of GLMM modelling of the effect of the speaker group on the variability (IQR) of the amount of tongue movement over syllable.

Fixed effect	Speakers	Estimate	MCMC mean	HPD 95% lower	HPD 95% upper	pMCMC
Number of lingual onset segments	AD TDC CAS	0.0006 -0.0005 -0.0026	0.0006 -0.0006 -0.0027	-0.0145 -0.0134 -0.0115	0.0156 0.0128 0.0061	0.9220 0.9174 0.4982
Number of non-lingual onset segments	AD TDC CAS	-0.0030 0.0045 -0.0042	-0.0029 0.0045 -0.0042	-0.0235 -0.0135 -0.0156	0.0188 0.0232 0.0081	0.7644 0.5788 0.4596

Table 24: Results of GLMM modelling of the effect of the number of lingual and non-lingual syllable onset segments on the rate of tongue movement over the syllable in each of the speaker groups.

Speakers	Syllable	Estimate	MCMC mean	HPD 95% lower	HPD 95% upper	pMCMC
AD vs. TDC	pay	-0.0203	-0.0203	-0.0296	-0.0108	0.0016
	say	-0.00292	-0.0291	-0.0370	-0.0212	0.0002
	lay	-0.0405	-0.0405	-0.0541	-0.0258	0.0002
	play	-0.0258	-0.0258	-0.0357	-0.0154	0.0010
	slay	-0.0278	-0.0278	-0.0391	-0.0160	0.0016
	splay	-0.0280	-0.0279	-0.0393	-0.0171	0.0001
AD vs. CAS	pay	-0.0214	-0.0216	-0.0353	-0.0078	0.0048
	say	-0.0145	-0.0145	-0.0260	-0.0029	0.0168
	lay	-0.0344	-0.0342	-0.0541	-0.0133	0.0010
	play	-0.0266	-0.0265	-0.0419	-0.0118	0.0010
	slay	-0.0270	-0.0271	-0.0442	-0.0099	0.0034
	splay	-0.0283	-0.0283	-0.0454	-0.0118	0.0014
TDC vs. CAS	pay	-0.0011	-0.0010	-0.0141	0.0139	0.8896
	say	0.0147	0.0148	0.0034	0.0262	0.0116
	lay	0.0061	0.0060	-0.0137	0.0273	0.5650
	play	-0.0008	-0.0008	-0.0163	0.0135	0.9224
	slay	0.0008	0.0009	-0.0165	0.0170	0.9260
	splay	-0.0004	-0.0003	-0.0168	0.0163	0.9792

Table 25: Results of GLMM modelling of the effect of the speaker group on the rate of tongue movement over the syllable for each of the syllable types.

Speakers	Syllable	Estimate	MCMC mean	HPD 95% lower	HPD 95% upper	pMCMC
AD vs. TDC	pay	-0.0099	-0.0099	-0.0204	0.0012	0.0732
	say	0.0007	0.0006	-0.0147	0.0173	0.9284
	lay	0.0039	0.0038	-0.0191	0.0249	0.7172
	play	-0.0029	-0.0028	-0.0148	0.0095	0.6380
	slay	-0.0084	-0.0084	-0.0273	0.0112	0.3614
	splay	-0.0030	-0.0031	-0.0218	0.0140	0.7376
AD vs. CAS	pay	-0.0170	-0.0170	-0.0325	-0.0012	0.0000
	say	0.0039	0.0041	-0.0195	0.0287	0.7182
	lay	-0.0080	-0.0083	-0.0413	0.0231	0.6052
	play	-0.0043	-0.0043	-0.0222	0.0137	0.6272
	slay	-0.0105	-0.0105	-0.0398	0.0157	0.4442
	splay	-0.0195	-0.0195	-0.0468	0.0056	0.1372
TDC vs. CAS	pay	-0.0072	-0.0072	-0.0225	0.0089	0.3578
	say	0.0032	0.0032	-0.0216	0.0271	0.7804
	lay	-0.0119	-0.0120	-0.0442	0.0213	0.4552
	play	-0.0014	-0.0014	-0.0192	0.0169	0.8720
	slay	-0.0021	-0.0021	-0.0309	0.0258	0.8918
	splay	-0.0165	-0.0168	-0.0423	0.0094	0.1988

Table 26: Results of GLMM modelling of the effect of the speaker group on the variability (IQR) of the rate of tongue movement over the syllable.

Fixed effect	Speakers	Estimate	MCMC mean	HPD 95% lower	HPD 95% upper	pMCMC
Number of onset segments	CAS1	1.8750	1.8740	1.2080	2.6070	0.0002
	CAS2	2.0150	2.0110	1.2910	2.7220	0.0001
	CAS3	1.6200	1.6180	0.6388	2.5240	0.0026

Table 27: Results of GLMM modelling of the effect of the number of syllable onset segments on syllable durations for each of the CAS speakers.

Fixed effect	Speakers	Estimate	MCMC mean	HPD 95% lower	HPD 95% upper	pMCMC
Number of lingual onset segments	CAS1 CAS2 CAS3	0.1621 0.1044 0.2866	0.1624 0.1046 0.2869	0.0203 -0.0290 0.1522	0.0210 0.2336 0.4241	0.0348 0.1034 0.0080

Table 28: Results of GLMM modelling of the effect of the number of lingual syllable onset segments on the amount of tongue movement over the syllable for each of the CAS speakers.

Fixed effect	Speakers	Estimate	MCMC mean	HPD 95% lower	HPD 95% upper	pMCMC
Number of lingual onset segments	CAS1 CAS2 CAS3	-0.0037 -0.0066 0.0024	-0.0038 -0.0667 0.0023	-0.0140 -0.0139 -0.0073	0.0073 0.0008 0.0122	0.4230 0.1825 0.5972
Number of non-lingual onset segments	CAS1 CAS2 CAS3	-0.0021 -0.0046 -0.0059	-0.0021 -0.0046 -0.0059	-0.0162 -0.0144 -0.0191	-0.0119 0.0055 0.0067	0.7244 0.4990 0.3354

Table 29: Results of GLMM modelling of the effect of the number of lingual and non-lingual syllable onset segments on the rate of tongue movement over the syllable for each of the CAS speakers.

Speakers	Segment	Estimate	MCMC mean	HPD 95% lower	HPD 95% upper	pMCMC
CAS1	/p/	0.4519	0.4681	-2.0000	2.7360	0.6462
	/s/	1.2290	1.2270	-0.8215	3.4360	0.2154
	/l/	3.8830	3.8790	1.5320	6.1550	0.0052
CAS2	/p/	1.0460	1.0650	-1.3490	3.3100	0.3036
	/s/	0.9334	0.9136	-1.4920	3.6010	0.4232
	/l/	2.8330	2.8360	-0.5932	6.5930	0.1038
CAS3	/p/	2.1100	2.0890	-0.7225	5.1170	0.1380
	/s/	0.8827	0.8989	-1.3380	3.1380	0.3414
	/l/	3.0550	3.0580	0.3392	5.7510	0.0342

Table 30: Results of GLMM modelling the effect of the type of syllable onset (single segment or cluster) on segment durations for each of the CAS speakers.

Appendix III. Patterns of the highest point of the tongue

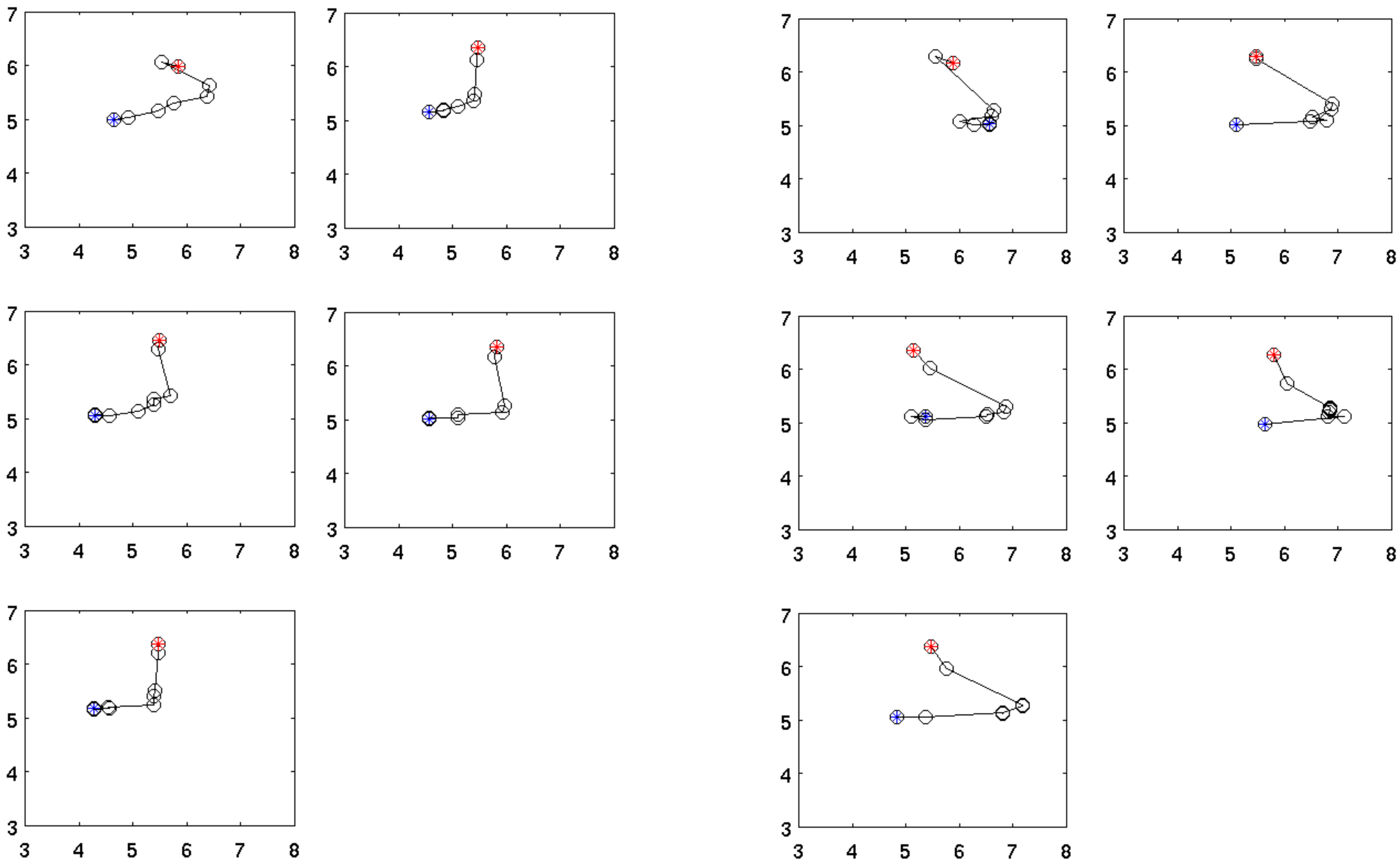


Figure 1: Speaker AD3's highest points on the midsagittal tongue contours over “pay” (left) and “say” (right). Scale is in cm.

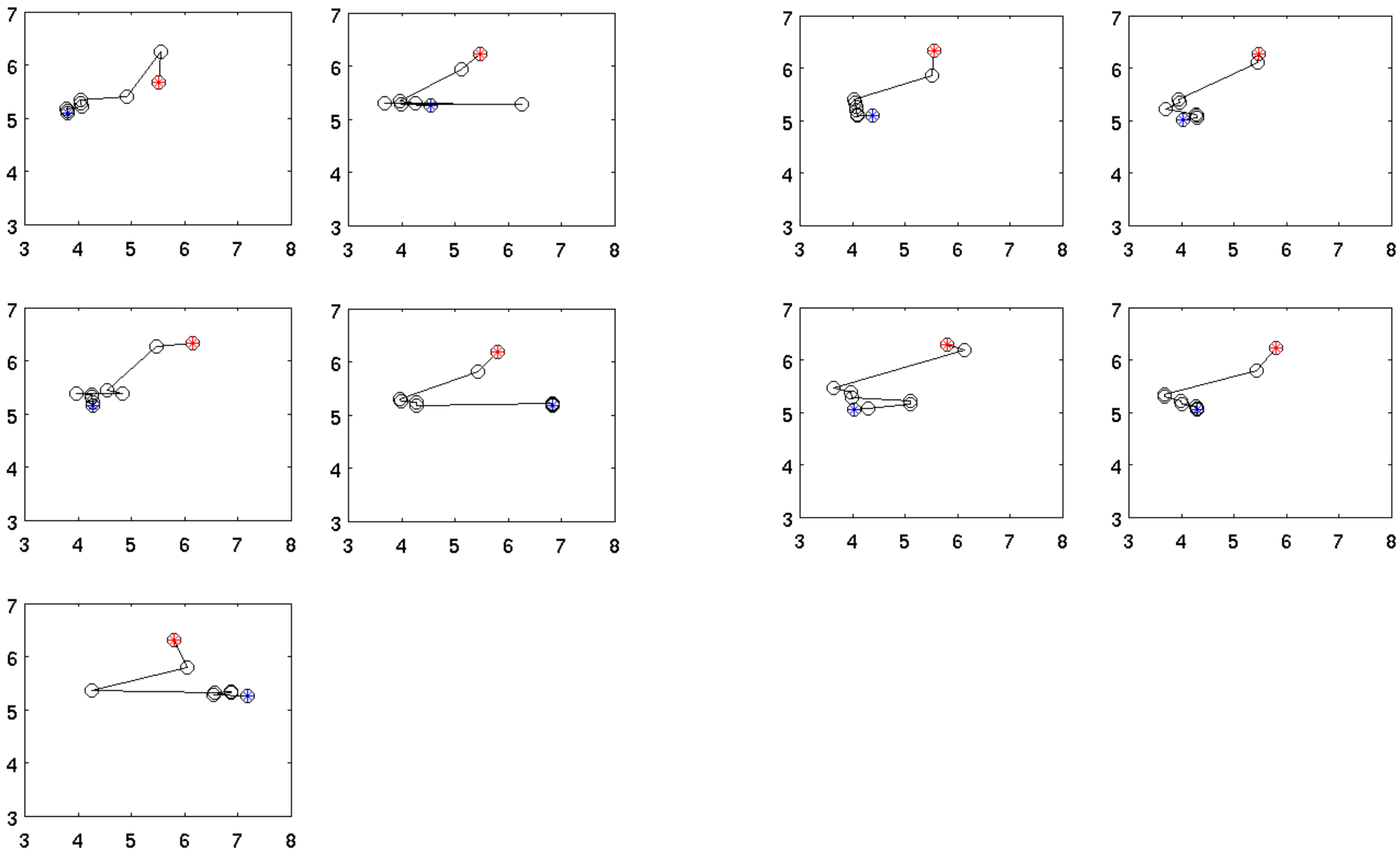


Figure 2: Speaker AD3's highest points on the midsagittal tongue contours over “lay” (left) and “play” (right). Scale is in cm.

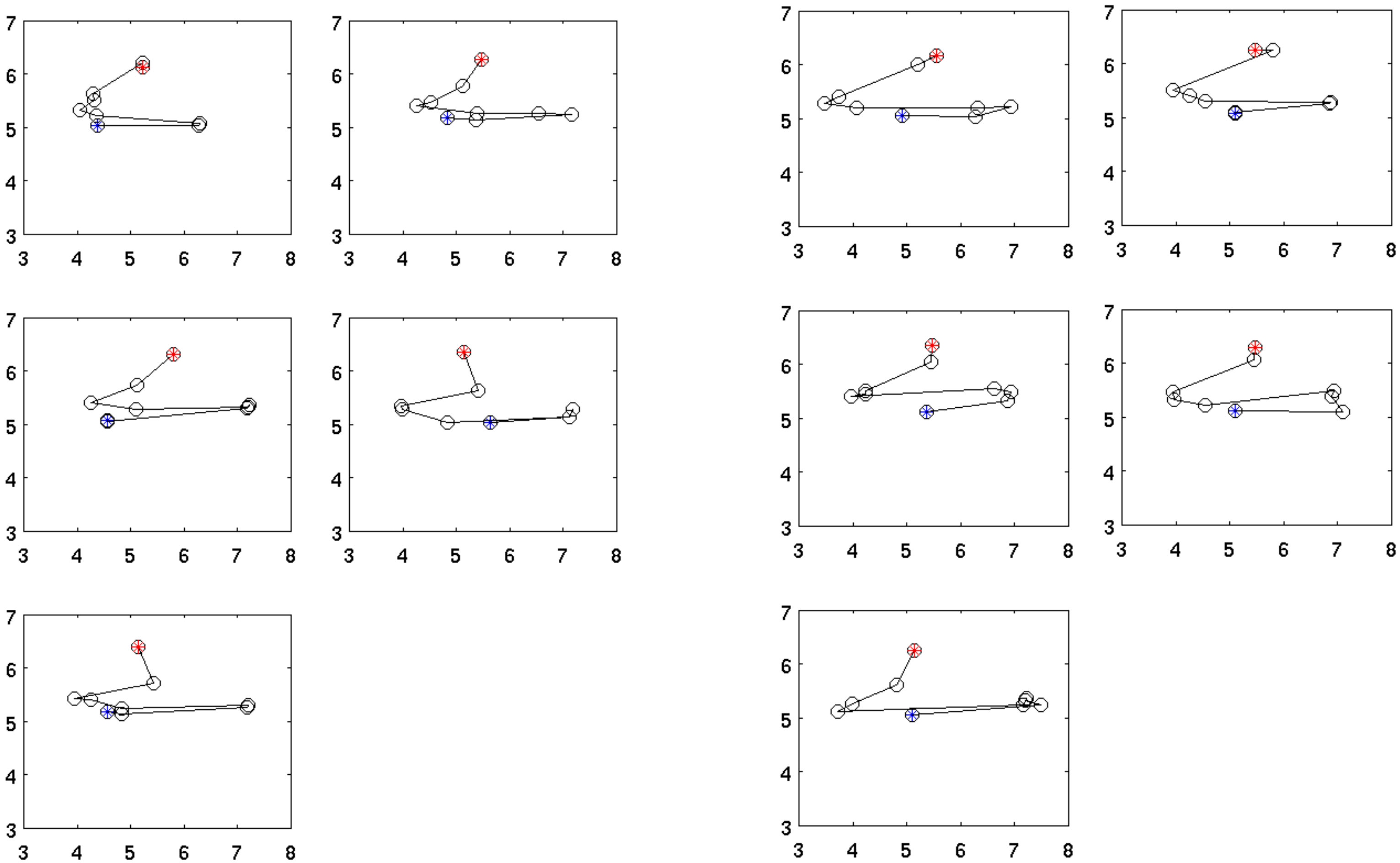


Figure 3: Speaker AD3's highest points on the midsagittal tongue contours over “slay” (left) and “splay” (right). Scale is in cm.

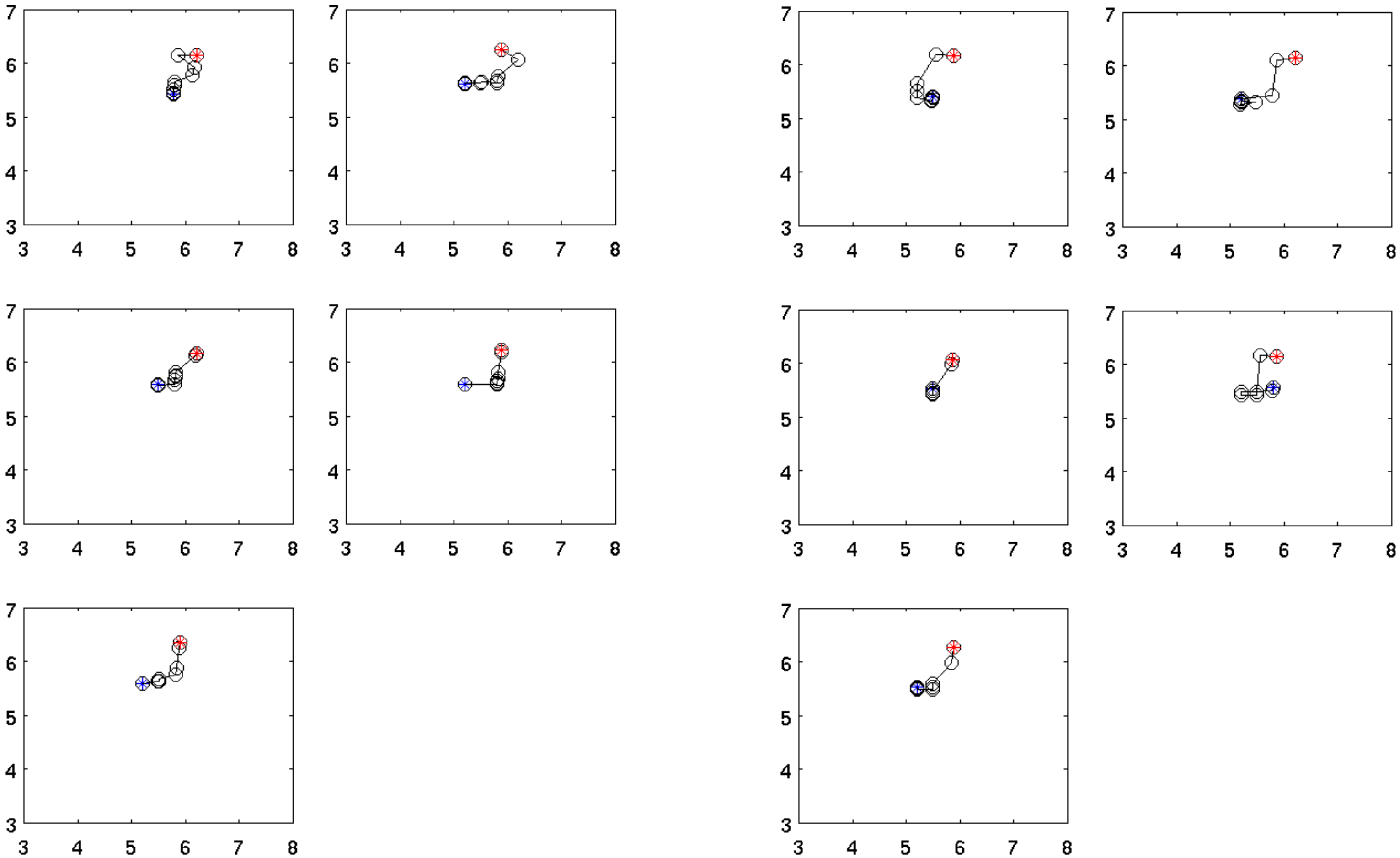


Figure 4: Speaker AD9’s highest points on the midsagittal tongue contours over “pay” (left) and “say” (right). Scale is in cm.

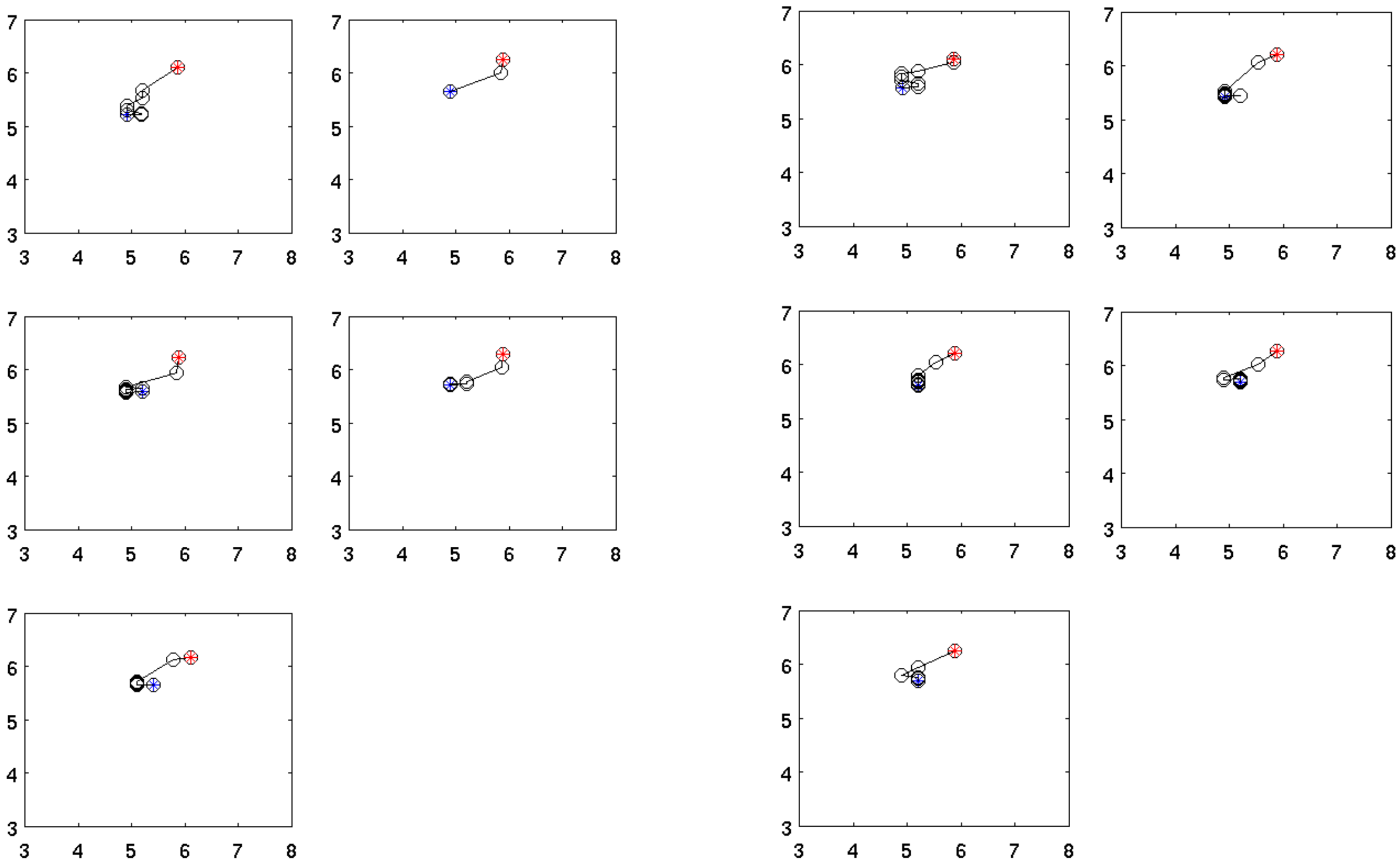


Figure 5: Speaker AD9's highest points on the midsagittal tongue contours over “lay” (left) and “play” (right). Scale is in cm.

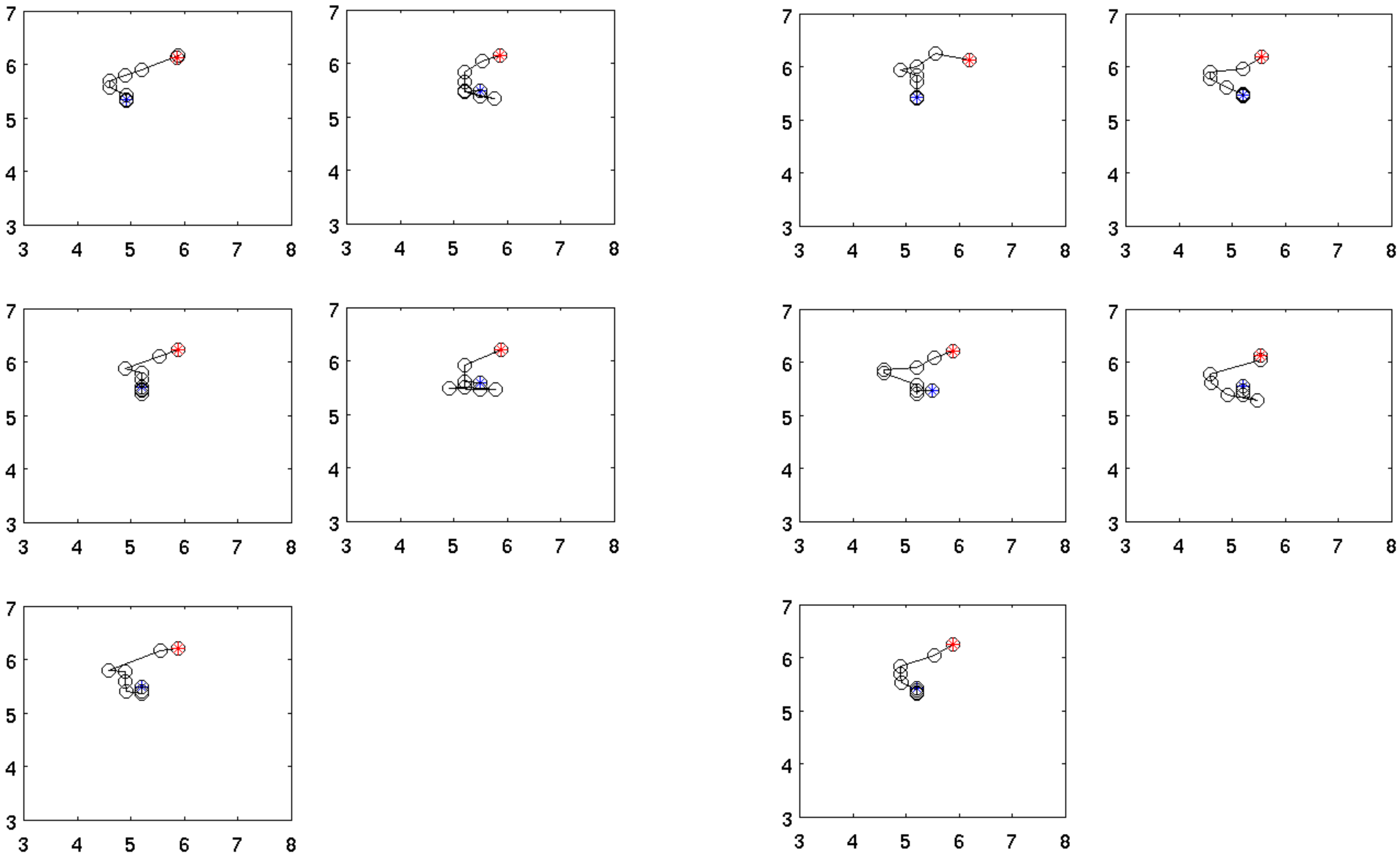


Figure 6: Speaker AD9’s highest points on the midsagittal tongue contours over “slay” (left) and “splay” (right). Scale is in cm.

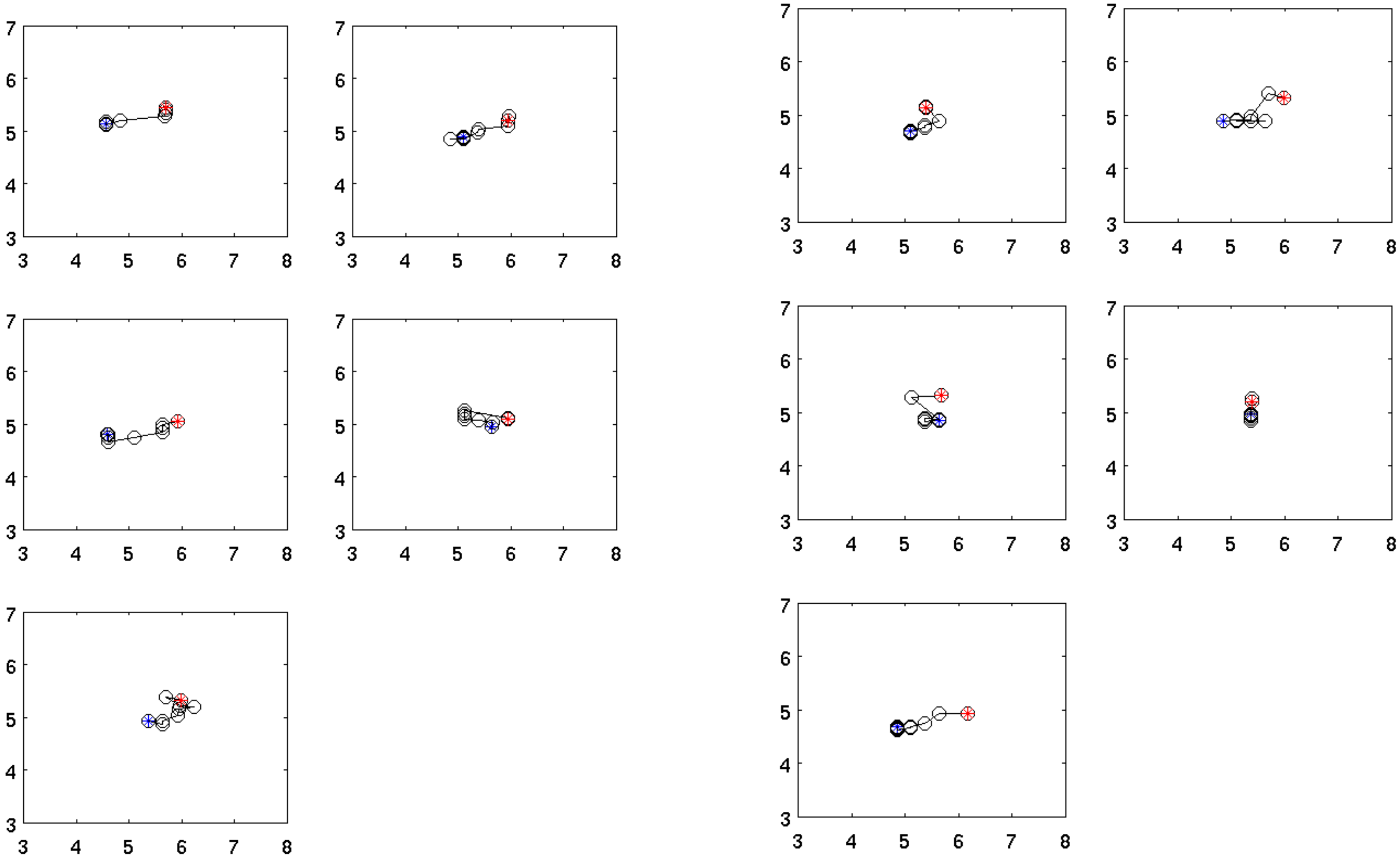


Figure 7: Speaker TDC9's highest points on the midsagittal tongue contours over "pay" (left) and "say" (right). Scale is in cm.

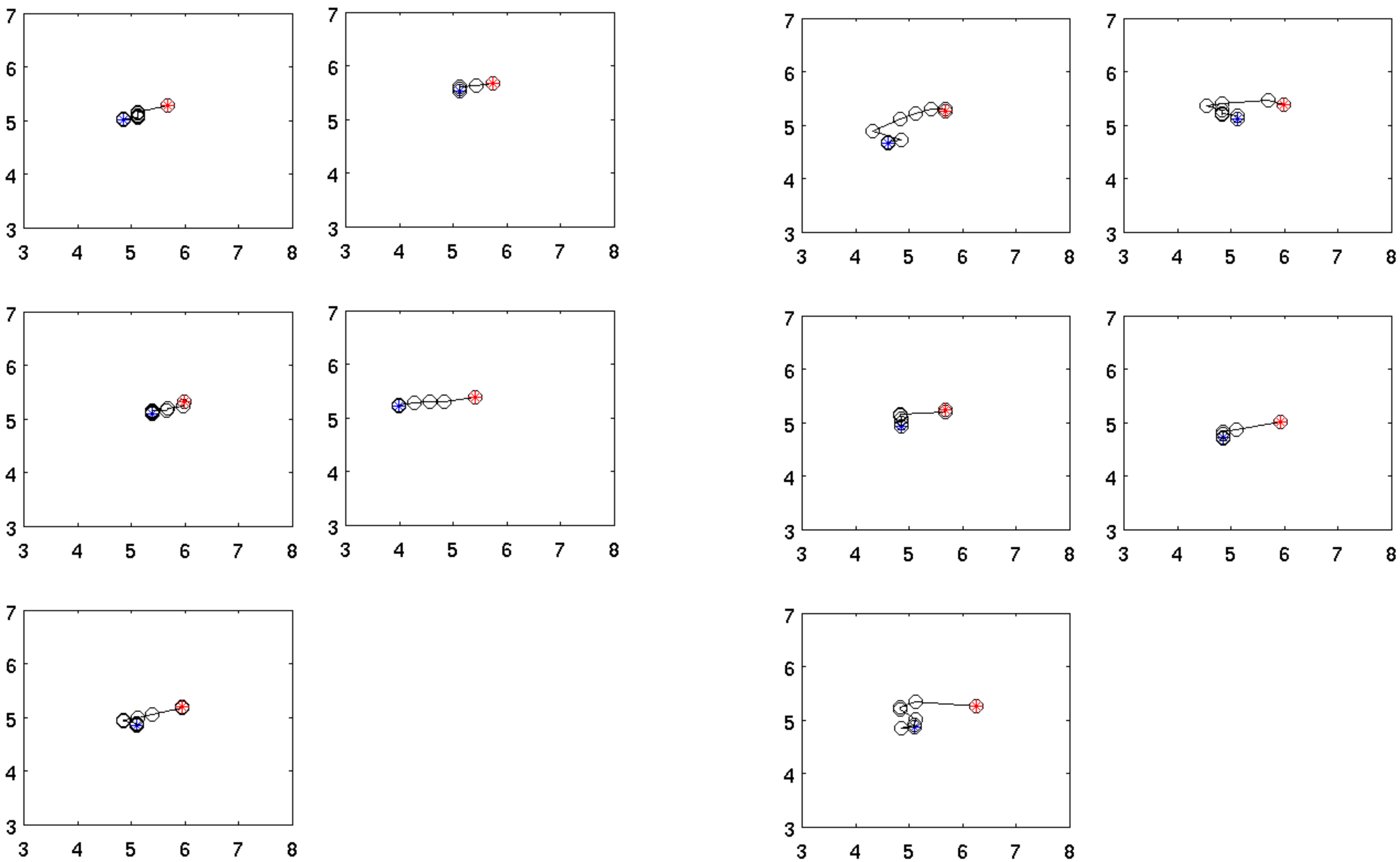


Figure 8: Speaker TDC9's highest points on the midsagittal tongue contours over "lay" (left) and "play" (right). Scale is in cm.

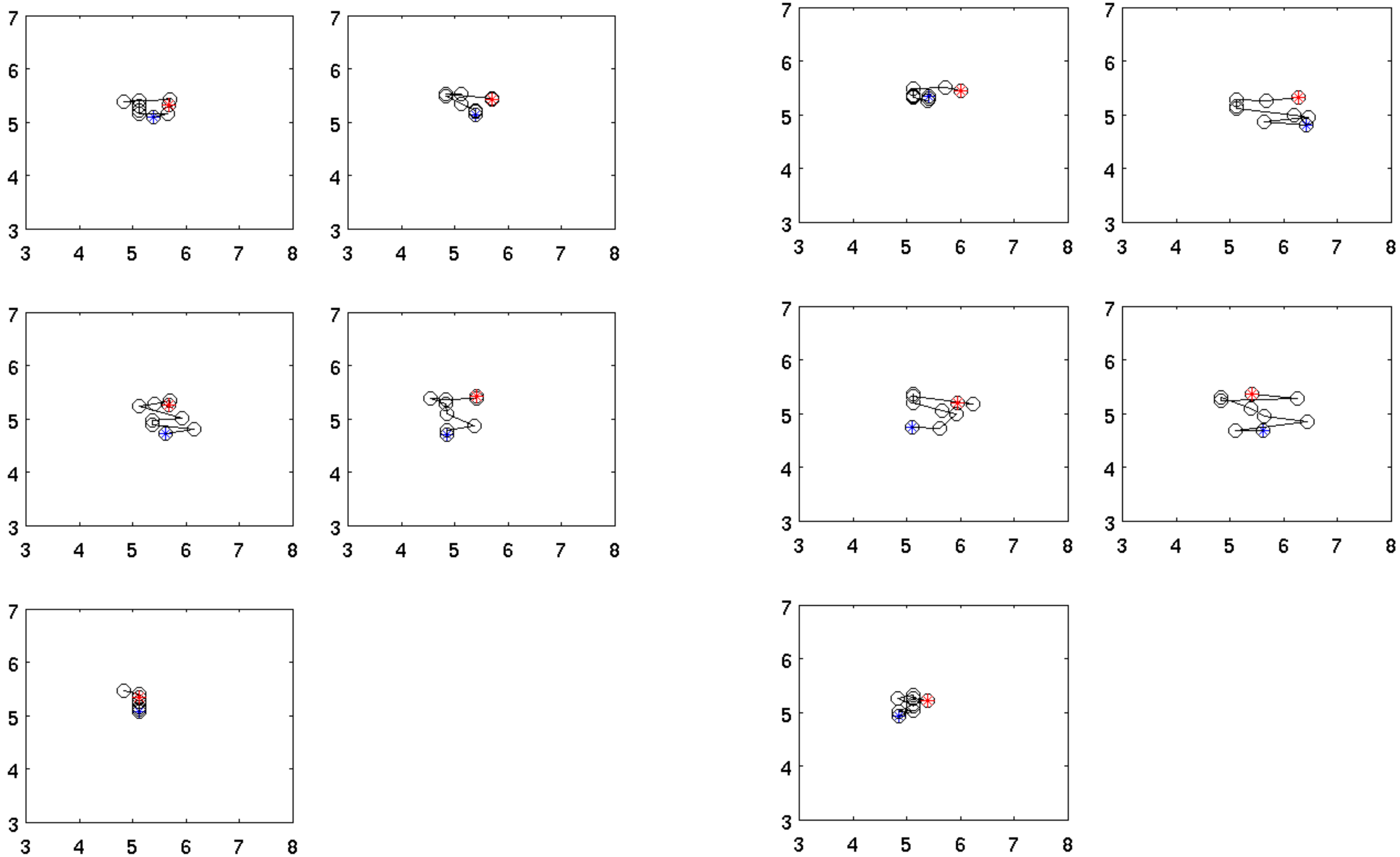


Figure 9: Speaker TDC9's highest points on the midsagittal tongue contours over “slay” (left) and “splay” (right). Scale is in cm.

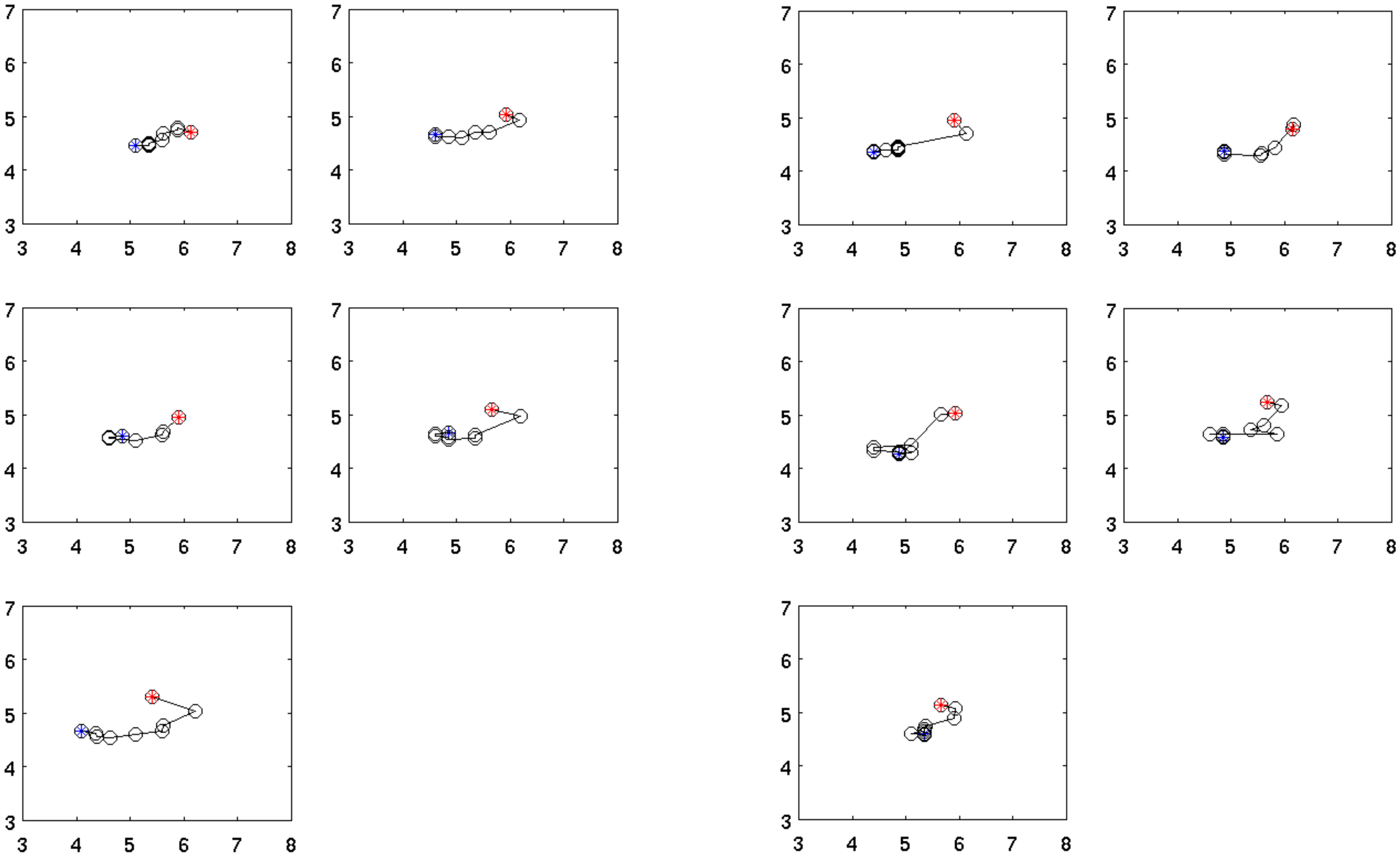


Figure 10: Speaker TDC10’s highest points on the midsagittal tongue contours over “pay” (left) and “say” (right). Scale is in cm.

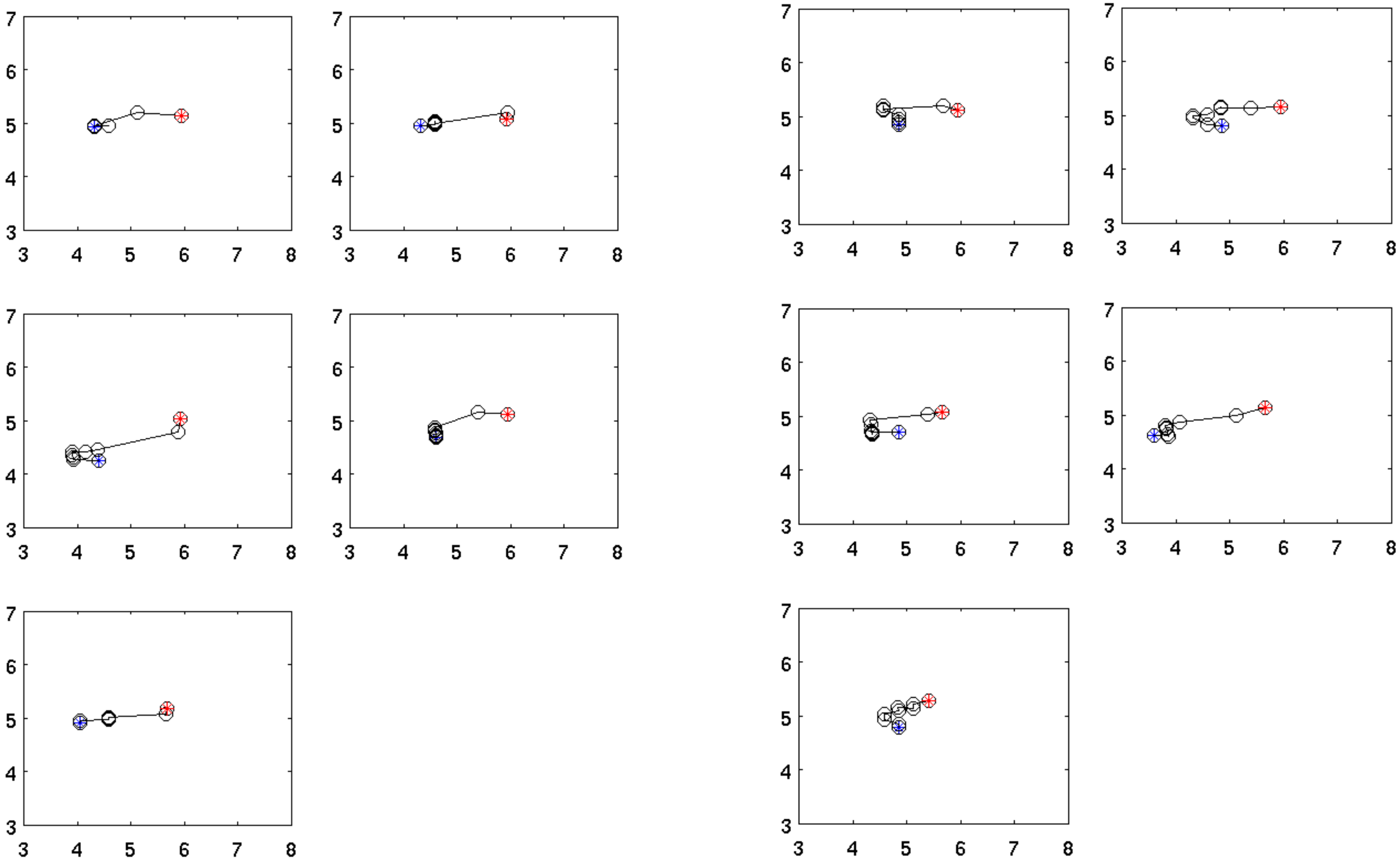


Figure 11: Speaker TDC10’s highest points on the midsagittal tongue contours over “lay” (left) and “play” (right). Scale is in cm.

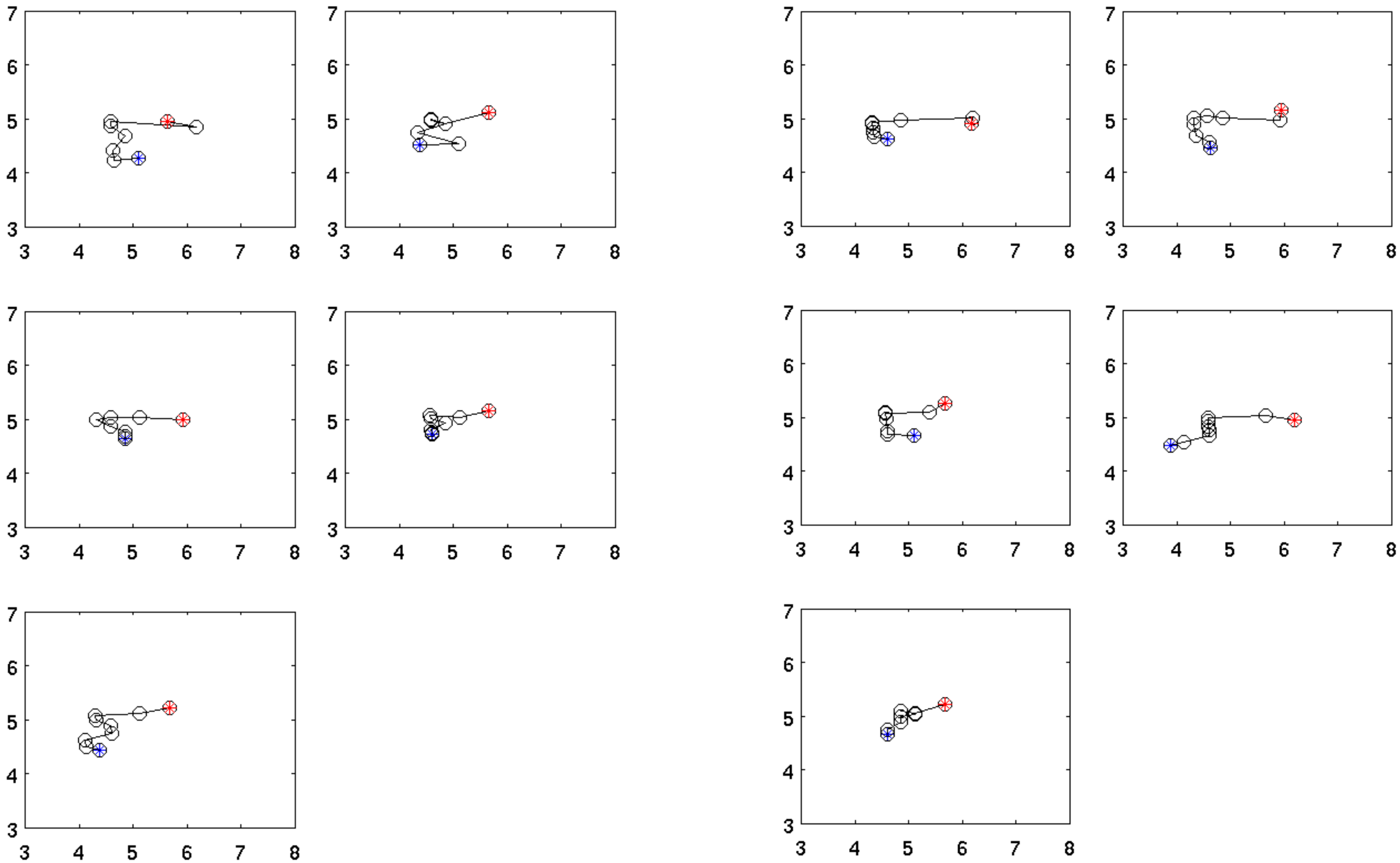


Figure 12: Speaker TDC10’s highest points on the midsagittal tongue contours over “slay” (left) and “splay” (right). Scale is in cm.

Appendix IV. Copies of published material

Tongue movements and syllable onset complexity: Ultrasound study

Tanja Kocjancic[1]
[1]Speech and Hearing Sciences, Queen Margaret University, UK
[1]Centre for Speech Technology Research, University of Edinburgh, UK

Abstract

In this study ultrasound was used to investigate tongue movements in syllables with different number and type of onset consonants. Ultrasound recordings provided the information of the distance the tongue travels over a target, and audio recordings of the time needed. The speed of the tongue's travel was calculated from the two measurements. Results of ten speakers have shown that both duration and distance travelled increase with an increased number of onset segments, but that distance travelled is additionally influenced by the type of the segment, as is speed. Duration also seemed to be the least speaker-dependant of the three parameters.

Key words: tongue movement, ultrasound, syllable onset.

1. Introduction

Ultrasound is a safe and non-invasive method which enables visualisation of tongue movement by depicting the upper tongue surface (Stone 2005). To obtain an ultrasound image of the tongue, the probe is placed under the speaker's chin and the inside of the mouth is scanned by emitting high frequency ultrasound waves. These waves are reflected at a boundary between mediums of different density (tongue/air and tongue/bone) and detected by the probe. After the detection the point of reflection is calculated and an image is created at that point.

Characteristics of differently structured syllable onsets have been investigated in several studies (Greenberg et al. 2003; Crystal and House 1990). The findings generally suggest that syllable duration is mainly affected by the syllable's stress and the number of segments. Syllable duration increases with an increasing number of segments in both stressed and unstressed condition. Stressed syllables and their segments have greater duration than their unstressed counterparts.

The aim of this study was to investigate how tongue movements are affected by changing the type and number of syllable onset segments. Tongue movements are described by the distance travelled by the tongue over an utterance, the duration of an utterance, and the tongue speed. It was hypothesized that both measurements will increase with the increasing number of onset segments, and that speed will vary, depending on the type of segments.

Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics 2008, 28-30 August 2006, Athens, Greece.

2. Methodology

Ultrasound data of ten native English female speakers, aged between 19 and 30 years, was analysed in this study. Speech material consisted of six monosyllabic real English words: pay, say, lay, play, slay, splay.

Midsagittal view of the tongue was recorded with Concept M6 (Dynamic Imaging) ultrasound with a frame rate of 30fps, and a special helmet was used to fix the probe under the speaker's chin. Participants repeated each of the words five times in a frame sentence "a [word] today". Both ultrasound and audio signals were recorded at the same time using Articulate Assistant Advanced, which allows temporal synchronisation of the two signals.

Articulate Assistant Advanced was used for annotating and tracing tongue contours on the recorded ultrasound images. All the reported results are for the "a [word]" part of the recording. The travelled distance of a target utterance was calculated as the sum of average nearest neighbour distances (aNND) between every pair of consecutive tongue contours of the utterance. aNND is an average of all the nearest neighbour distances measured between the points on the two contours of a pair. Duration was measured from audio signal, and the tongue speed was calculated as the distance travelled over duration, to give information about the relationship between the two measurements.

3. Results

3.1. Duration

The results showed that duration increases with the increasing number of syllable segments (Figure 1a) and that it is statistically significant between all pairs of targets (Figure 1d, solid lines). Additionally, most of the speakers showed the same pattern of increasing duration: single onsets shorter than two consonants clusters, which were also shorter than the three consonants one.

3.2. Distance travelled by the tongue

Overall, the distance travelled by the tongue did increase with the increased number of onset segments (Figure 1b) but the increase was not stat. sig. between all of the syllable pairs (Figure 1d, dashed lines). Distribution of the measurements was also more variable than in the case of duration, and individual speakers showed less similar patterns of increasing the distance travelled by the tongue over a syllable. Not all speakers had greater measurements for targets with clustered onset than for those with single onset.

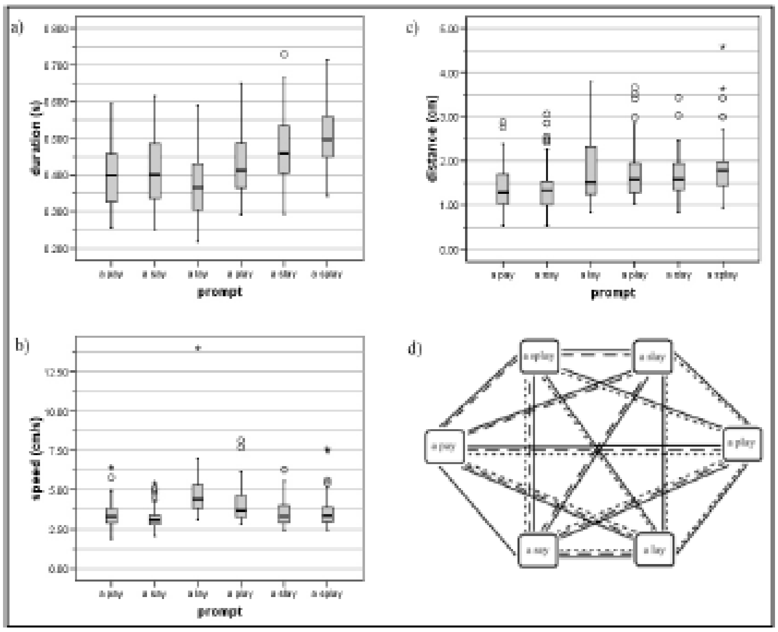


Figure 1. Duration (a), tongue's distance travelled (b) tongue speed (c) and stat. sig. (Wilcoxon Signed Rank Test, $p < 0.05$) between pairs of targets (d; dashed line = distance travelled, solid line = duration, dotted line = tongue speed.)

3.3. Speed of the tongue's travelling

Speed of the tongue's travelling is presented in Figure 1c. The tongue travelled the fastest over "a lay" and the slowest over "a say" with most of the target pairs being stat. sig. different from each other (Figure 1d, dotted lines). Number of onset segments did not influence tongue speed, and individual speakers showed very different patterns of increasing tongue speed over targets, although eight out of ten speakers had the fastest tongue speed over "a lay".

4. Discussion and conclusion

Analysis has shown that duration increases with the addition of segments to syllable onset and thus confirmed the results of previous studies (Greenberg et al. 2003; Crystal and House 1990). Results of distance travelled measurements also showed the influence of the number of onset segments, as single onset targets were shorter than clustered ones. However, they were also influenced by the type of the segment. As expected, targets with single onset

/p/ and /s/ were shorter than those with single /l/, which had a distance travelled more similar to the targets with clustered onsets. /p/ is not a lingual consonant and does not contribute any tongue movement to the distance travelled, and /s/ has less tongue movement than /l/. In contrast, /p/ does contribute to the distance travelled in case of clustered onsets as it seems to make movement for /l/ more prominent and less restricted than in case of /sl/ clusters. Tongue speed, on the other hand, does not seem to depend on the number of segments, but mainly on the type.

Based on the results of this study it can also be concluded that duration of spoken target is less speaker-dependant than distance travelled. The latter depends on the size of individual speech organs and can not be adapted as duration can. Consequently, tongue speed is the most speaker-dependant and is appropriately adapted depending on the demands of the space over which the tongue has to travel, and the demands of appropriate speech timing.

This study has demonstrated that ultrasound is sensitive enough to describe continuous tongue movement during speech although it has some limitations. The most important is that it does not produce an image of raised tongue tip with air pocket below it, and thus we miss information about that part of the tongue. Measurements could be also affected by the probe which could potentially restrict jaw movement. Mooshammer et al. (2003) have shown that jaw position is low for /l/ and high for /s/. However, this effect is expected to affect all participants.

Acknowledgements

Many thanks to my supervisor, Dr Nigel Hewlett. Financial support provided by the Marie Curie EdSST programme.

References

Articulate Instruments Ltd. 2007. Articulate Assistant Advanced User Guide: Version 2.07. Edinburgh, UK: Articulate Instruments Ltd.

Crystal, T. H. and House, A. S. 1990. Articulation rate and the duration of syllables and stress groups in connected speech. Journal of the Acoustical Society of America 88 , 101-112.

Greenberg, S., Carvey, H., Hitchcock, L. and Chang, S. Y. 2003. Temporal properties of spontaneous speech - a syllable-centric perspective. Journal of Phonetics 31, 465-485.

Mooshammer, C., Geumann, A., Hoole, P, Alfonso, P., van Lieshout, P. and Fuchs, S. 2003. Coordination of lingual and mandibular gestures for different manners of articulation. In Proceedings of the 15th ICPhS, 81-84, Barcelona, Spain.

Stone, M. 2005. A guide to analysing tongue motion from ultrasound images. Clinical Linguistics & Phonetics 19, 455-501.

Ultrasound investigation of tongue movements in syllables with different onset structure

Tanja Kocjančič
Queen Margaret University
E-mail: tkocjancic@qmu.ac.uk

Abstract

This study is an attempt to describe syllables with different onset structure not only in terms of durational changes but using ultrasound also in terms of the distance the tongue travels over a syllable by using ultrasound and to compare the ratio between the two parameters, expressed as speed. Results indicate that both measures increase with an increasing number of onset segments but not to the same degree for all targets. Therefore speed was not constant over all of them. Additionally, type of onset constituent greatly influenced the three parameters and there were large between-speaker similarities in case of durational changes.

1 Introduction

Ultrasound is a safe and non-invasive imaging technique that enables visualisation of the tongue inside the mouth during speech without placing any obstructions on the tongue. Ultrasound images are recorded by placing the probe under the speaker's chin. The probe emits ultrasound waves which are reflected at a boundary between two media of different density (i.e. tongue/air or tongue/bone boundary). Based on the time lapse between incident and reflected wave, a point of reflection is calculated and an image is created at that point. Since the reflection at the air boundary is almost 100 % it cannot image a raised tongue tip with an air pocket below it. Nevertheless, ultrasound allows observation of tongue during speech and therefore provides a good tool for investigating most aspects of tongue movement.

Properties of differently structured syllables have been investigated in several studies and one of the general conclusions is that syllable duration increases with an increasing number of onset segments [5, 2]. The increase in duration, however, is not simply a sum of the individual durations of constituent segments. Due to coarticulation individual segments influence each other's articulatory realisation and consequently duration. Most of the studies report shortening of segments in clusters. However, only average group results and not measurements of individual segments are usually reported. Detailed inspection reveals that segments in cluster onsets can be either shorter, longer or the same as in single onset position depending on the number of the onset segments, the position within onset, a segment's intrinsic duration and the identity of adjacent segments. Since the onsets in the present study are composed of the consonants /p/, /s/, and /l/, some past findings regarding these segments are described.

It has been observed that in English a singleton syllable-initial /s/ is the longest of the three consonants, followed by /p/, and then by /l/, which is the shortest [8]. When these segments are part of an onset cluster they each change in their own specific way. Initial /s/ is shorter when followed by a stop [8, 7] and longer [5], equal [2] or shorter [8] when followed by /l/ [5, 6, 9]. Shortening of initial /s/ when followed by a consonant was also found to occur in Italian [4].

English /p/ in an initial onset position lengthens when followed by a voiced consonant [8, 9], and shortens as a second segment in a cluster [8, 7]. In Italian however, the duration of stop consonants is not affected by the onset size or by the position within the onset [4].

In contrast to the other two consonants /l/ can not be the first segment of a consonant cluster in English. In the second position of a cluster /l/ shortens when preceded by a fricative [8, 6, 9] and either shortens [8] or stays the same when it follows an unvoiced stop [9].

The three segments do not differ only in their durations but also in their articulation. The relationship between these two parameters is not completely clear. O'Shaughnessy [8] stated that shorter durations of segments in a cluster result from shorter distances the articulators have to travel in the realisation of a cluster. Umeda [9], on the other hand, relates the duration of a consonant to the articulator and/or type of gesture shared between the consonant and its adjacent consonant. Duration of the consonant is different when the gesture is overlapping than when it is conflicting. Sharing a gesture additionally prevents consonant's shortening, and sharing an articulator (e.g. tongue in /st/) lowers the variance of timing. O'Shaughnessy [8] reported that in French consonants were shorter in a cluster of two segments sharing place of articulation and longer when their place of articulation differed. The same was not observed in a study of English [1].

The main difference between the segments /p/, /s/ and /l/, from the point of tongue investigation, is that /p/ is a non-lingual consonant while /s/ and /l/ are both lingual. As such, /p/ does not have any tongue movement necessary for its realisation but the other two require the correct tongue position. Fricatives have to restrict the tongue dorsum position to achieve the necessary tongue constriction while articulation of laterals is less constrained [3].

The aim of this study was to investigate how tongue movements are affected by changing the type and number of syllable onset segments. Tongue movements are described by the tongue's distance of travel over an utterance, the duration of an utterance and the ratio between the two measurements, expressed as speed of the tongue's distance of travel. The following hypotheses where tested in this study: (i) the addition of either a lingual or non-lingual consonant to the syllable onset increases the duration of the syllable, (ii) the addition of a lingual consonant increases the distance the tongue travels over a syllable while the addition of non-lingual consonant does not. Therefore, for example, /a lay/ will have shorter duration than /a play/ but they will have similar distance, resulting in higher distance to duration ratio (speed) for /a lay/ than for /a play/.

2 Methodology

2.1 Speakers

Ultrasound data from 10 native English female speakers, aged between 19 and 30 years, was analysed in this study. Speech material consisted of six mono-syllabic real English words: "pay", "say", "lay", "play", "slay", "splay".

2.2 Speech material and recording

A midsagittal view of the tongue was recorded with Concept M6 (Dynamic Imaging) ultrasound with a frame rate of 30fps. A special helmet was used to fix the probe under the speaker's chin. Participants repeated each of the words five times in a frame sentence "a [word] today". Both ultrasound and audio signals were recorded at the same time using Articulate Assistant Advanced which allows temporal synchronisation of the two signals.

2.3 Data analysis

Articulate Assistant Advanced was used for annotating and tracing tongue contours on the recorded ultrasound images. All the reported results are for the "a [word]" part of the recording. The distance of tongue travel in a target utterance was calculated as a sum of average nearest neighbour distances (aNND) between every pair of consecutive tongue contours of the utterance. aNND is an average of all the nearest neighbour distances measured between the points on the two contours of a pair. Duration was measured from audio signal and speed was calculated as the ratio of distance and duration. Stat. sig. was tested using a non-parametric Friedman's ANOVA with a post-hoc Wilcoxon signed-rank test and Bonferroni correction.

3 Results and discussion

Group results for the ten speakers are presented in Figure 1. As can be seen in Figure 1a, the duration of the target increases with increasing number of onset segments and the target with single onset /s/ is the longest of the single onset ones and the one with /l/ the shortest. These two observations support findings of previous studies. Additionally, duration

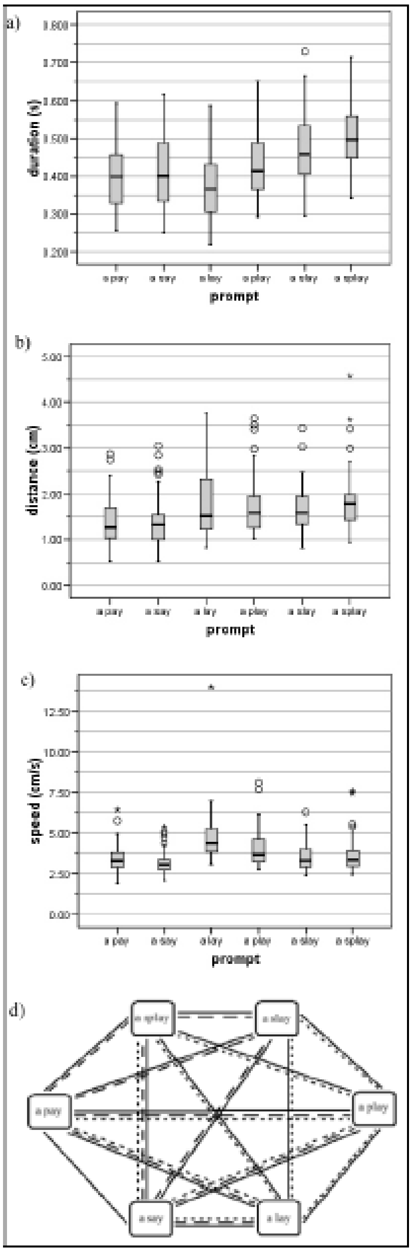


Figure 1: *Duration (a), distance of tongue's travel (b), speed of tongue's distance of travel over the target (c) and stat. sig. (p <0.0033) between pairs of targets (d; dashed line = tongue's distance of travel, solid line = duration, dotted line = speed.)*

is significantly different between all targets except "a pay"-"a say" and "a say"-"a play" pairs (solid lines in Figure 1d; lines between the pairs of targets represent stat. sig. difference) showing a strong effect of the type of onset segments on the total duration. However, despite the difference there is a similarity in the distribution of measurements of the six targets as they all show similar variation. Moreover, similarities were observed between individual speakers as well. They showed very similar patterns of increasing duration, with nine out of ten having the shortest duration for "a lay" and longest for "a splay".

The second measured parameter, distance of travel, has revealed the same effect of the number of onset segments as duration; as the onset increases, the distance travelled increases as well (Figure 1b). The measurement was the shortest in "a pay" and "a say", which are not significantly different from each other but are significantly shorter than all the other targets (Figure 1d). This result was expected for "a pay" since /p/ is not a lingual segment but not for "a say". A possible explanation is that although /s/ has a lingual component in its realisation, the movement from the /ə/ with a tongue positioned more or less in the centre of the mouth, is very small. Additionally, this result could be affected by the limitation of ultrasound which does not image a raised tongue tip. The greatest movement could be done with the tip of the tongue, and not recorded, while the body stays in a position similar to /ə/. Another observation regarding this parameter is that "a lay" had similar distance travelled to targets with cluster onsets (Figure 1d). This is the result of the /l/ having the greatest lingual component of the three consonants and the others having less influence on the total distance travelled over targets when in clusters with /l/. In fact, /p/ should not have any influence and /s/ should show some. Following this, it was expected that the tongue would travel less distance over "a play" than over "a slay" but the results showed no difference. Further inspection of relevant ultrasound frames has revealed that in case of the /pl/ cluster the tongue movement for /l/ is not restricted by /p/ and can be thus fully realised. On the other hand, in case of /sl/, tongue movement for /s/ restricts movement for /l/ and gives the tongue less space to cover in the articulation of the later. Similarly, in the case of /spl/ the tongue articulates /s/, stays in the same position during /p/ and then realizes /l/, which is consequently again restricted by the last preceding lingual consonant (/s/). The result is a similar distance distance travelled over a /sl/ than over a /spl/ target (Figure 1d). A very notable difference between the six targets is the distribution of the measurements, with “a lay” being the most variable and “a splay” the least (based on interquartile range). This can be due to singleton /l/ being the least constrained in its articulation of the three consonants [4] and possibly the most constrained of all segments when in cluster. Another difference was observed between speakers when comparing the increasing pattern of distance travelled over the target. Individual speakers showed very different patterns with only one common factor; most of them had shorter travelled distances in targets with single onsets than in those with clusters. Such a result may be influenced by the differences in the size of the tongue and oral cavity between speakers.

As already seen, both duration and the distance travelled showed an effect of the increasing number of onset segments but their ratio, expressed as speed, did not (Figure 1c). Although the change of both parameters was in the same direction (increase) the extent of each of them was not the same for all targets. The exceptions were “a lay” and “a play”, which were significantly faster from almost all the others (similar speed over “a pay” and “a play”) with “a lay” being even faster than “a play” (Figure 1d). These results, however, simply reflect the measured changes in distance travelled and duration. Targets with shortest durations and longer distance of travel had higher distance to duration ratio and thus higher speed of tongue movement over the entire target. This parameter has also shown only one common characteristic between the ten speakers: speed was the highest in “a lay” for seven of them and the second highest for the remaining three.

4 Conclusion

The data has confirmed the expected result of duration increasing with increasing number of onset segments, either lingual or non-lingual, but not the expectation that the tongue’s distance of travel would increase only with the addition of a lingual consonant. Distance showed a greater effect of individual segments, e.g. /l/ having the greatest tongue movement of the three onset consonants as a singleton and the greatest effect on the tongue’s distance of travel over a cluster onset target. Overall, both measured parameters increased with the increasing number of onset segments but not always to the same degree causing their ratio, speed, to be constant for some targets but not for all, which reflects the characteristic of individual consonants and their combinations in clusters. Comparing individual speaker data revealed that duration was the least speaker dependant and had the most similar variability across targets of the three parameters.

5 Acknowledgements

Many thanks to my supervisor, Dr Nigel Hewlett. Financial support provided by the Marie Curie EdSST programme.

References

[1] T. Crystal and A. House. Segmental durations in connected-speech signals: Current results. *The Journal of the Acoustical Society of America*, 83(4):1553–1573, 1988.

[2] T. Crystal and A. House. Articulation rate and the duration of syllables and stress groups in connected speech. *The Journal of the Acoustical Society of America*, 88(1):101–112, 1990.

[3] E. Farnetani. *Coarticulation and connected speech processes*. Blackwell Publishers, 1997.

[4] E. Farnetani and S. Kori. Effects of syllable and word structure on segmental durations in spoken Italian. *Speech Communication*, 5(1):17–34, 1986.

[5] S. Greenberg, H. Carvey, L. Hitchcock, and S. Chang. Temporal properties of spontaneous speech-a syllable-centric perspective. *Journal of Phonetics*, 31(3):465–485, 2003.

[6] M. Haggard. Abbreviation of consonants in English pre-and post-vocalic clusters. *Journal of Phonetics*, 1(1):9–24, 1973.

[7] D. Klatt. The duration of [s] in English words. *Journal of Speech, Language and Hearing Research*, 17(1):51–63, 1974.

[8] D. O’Shaughnessy. Consonant durations in clusters. *IEEE Transactions on Acoustics Speech and Signal Processing*, 22(4):282–295, 1974.

[9] N. Umeda. Consonant duration in American English. *The Journal of the Acoustical Society of America*, 61(3):846–858, 1977.

References

American Speech-Language-Hearing Association (2007). Childhood apraxia of speech [technical report]. Retrieved 12 January 2008 from http://www.asha.org/policy.

Articulate Instruments Ltd. (2008). *Ultrasound stabilisation headset users manual: Revision 1.13*. Edinburgh.

Baayen, R. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press, Cambridge.

Baayen, R., Davidson, D., and Bates, D. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59:390–412.

Barry, R. (1993). Measuring segmental timing in pathological speech using electropalatography. *Clinical Linguistics & Phonetics*, 7:275–283.

Bernhardt, B., Gick, B., Bacsfalvi, P., and Adler-Bock, M. (2005). Ultrasound in speech therapy with adolescents and adults. *Clinical Linguistics & Phonetics*, 19:605–617.

Bernhardt, B., Gick, B., Bacsfalvi, P., and Ashdown, J. (2003). Speech habilitation of hard of hearing adolescents using electropalatography and ultrasound as evaluated by trained listeners. *Clinical Linguistics & Phonetics*, 17:199–216.

Betz, S. and Stoel-Gammon, C. (2005). Measuring articulatory error consistency in children with developmental apraxia of speech. *Clinical Linguistics & Phonetics*, 19:53–66.

Blackwell, R. (2001). Ultrasound equipment. In H.B. Meire & D.O. Cosgrove & K.C. Dewbury & P. Farrant (Eds.), *Clinical Ultrasound a comprehensive text. Abdominal and general ultrasound - Vol. 1* (pp.13–36). Churchill Livingstone, London.

Blech, A., Springer, L., and Kroger, B. (2007). Perceptual and acoustic analysis of vowel productions in words and pseudo words in children with suspected childhood apraxia of speech. In *International Congress of Phonetic Sciences, ICPhS*.

Boersma, P. and Weenink, D. (2008). Praat: Doing phonetics by computer [Computer program]. Version 5.0.22. Retrieved 27 April 2008 from http://www.praat.org/.

Boyce, S. and Schmidlin, S. (2009). Using ultrasound with therapy for resistant /r/. [presentation] November 21. New Orleans: 2009 ASHA Convention.

Cabin, R. and Mitchell, R. (2000). To Bonferroni or not to Bonferroni: When and how are the questions. *Bulletin of the Ecological Society of America*, 81:246–248.

Cheng, H., Murdoch, B., and Goozee, J. (2007a). Temporal features of articulation from childhood to adolescence: An electropalatographic investigation. *Clinical Linguistics & Phonetics*, 21:481–499.

Cheng, H., Murdoch, B., Goozee, J., and Scott, D. (2007b). Electropalatographic assesment of tongue-to-palate contact patterns and variability in children, adolescents, and adults. *Journal of Speech, Language and Hearing Research*, 50:375–392.

Cheng, H., Murdoch, B., Goozee, J., and Scott, D. (2007c). Physiologic development of tongue-jaw coordination from childhood to adulthood. *Journal of Speech, Language and Hearing Research*, 50:352–360.

Crystal, T. and House, A. (1982). Segmental durations in connected speech: Preliminary results. *Journal of the Acoustical Society of America*, 72:705–716.

Crystal, T. and House, A. (1988). Segmental durations in connected-speech signals: Current results. *Journal of the Acoustical Society of America*, 83:1553–1573.

Crystal, T. and House, A. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America*, 88:101–112.

Davis, B., Jacks, A., and Marquardt, T. (2005). Vowel patterns in developmental apraxia of speech: Three longitudinal case studies. *Clinical Linguistics & Phonetics*, 19:249–274.

Davis, B., Jakielski, K., and Marquardt, T. (1998). Developmental apraxia of speech: Determiners of differential diagnosis. *Clinical Linguistics & Phonetics*, 12:25–45.

DiSimoni, F. (1974a). Effect of vowel environment on the duration of consonants in the speech of three-, six-, and nine-year-old children. *Journal of the Acoustical Society of America*, 55:360–361.

DiSimoni, F. (1974b). Influence of consonant environment on the duration of vowels in the speech of three-, six-, and nine-year-old children. *Journal of the Acoustical Society of America*, 55:362–363.

Farnetani, E. and Kori, S. (1986). Effects of syllable and word structure on segmental durations in spoken Italian. *Speech Communication*, 5:17–34.

Field, A. (2009). *Discovering statistics using SPSS*. Sage, London.

Fisher, S., Vargha-Khadem, F., Watkins, K., Monaco, A., and Pembrey, M. (1998). Localisation of a gene implicated in a severe speech and language disorder. *Nature Genetics*, 18:168–170.

Flipsen Jr., P. and Gildersleeve Neumann, C. (2009). Childhood apraxia of speech: Some basics of assessment and treatment. [presentation] November 19. New Orleans: 2009 ASHA Convention.

Forrest, K. (2003). Diagnostic criteria of developmental apraxia of speech used by clinical speech-language pathologists. *American Journal of Speech-Language Pathology*, 12:376–380.

Gerosa, M., Lee, S., Giuliani, D., and Narayanan, S. (2006). Analyzing children's speech: An acoustic study of consonants and consonant-vowel transition. In *2006 IEEE International Conference on Acoustics, Speech, and Signal Processing*.

Gilbert, J. and Purves, B. (1977). Temporal constraints on consonant clusters in child speech production. *Journal of Child Language*, 4:417–432.

Goozee, J., Murdoch, B., Ozanne, A., Cheng, Y., Hill, A., and Gibbon, F. (2007). Lingual kinematics and coordination in speech-disordered children exhibiting differentiated versus undifferentiated gestures. *International Journal of Language Communication Disorders*, 42:703–724.

Green, J. and Wang, Y. (2003). Tongue-surface movement patterns during speech and swallowing. *Journal of the Acoustical Society of America*, 113:2820–2833.

Greenberg, S., Carvey, H., Hitchcock, L., and Chang, S. (2003). Temporal properties of spontaneous speech: A syllable-centric perspective. *Journal of Phonetics*, 31:465–485.

Groenen, P., Maassen, B., Crul, T., and Thoonen, G. (1996). The specific relation between perception and production errors for place of articulation in developmental apraxia of speech. *Journal of Speech and Hearing Research*, 39:468–482.

Haggard, M. (1973). Abbreviation of consonants in English pre- and post-vocalic clusters. *Journal of Phonetics*, 1:9–24.

Hangiandreou, N. (2003). AAPM/RSNA Physics tutorial for residents: Topics in US. M-mode US: Basic concepts and new technology. *RadioGraphics*, 23:1019–1033.

Iskarous, K. (2005). Patterns of tongue movement. *Journal of Phonetics*, 33:363–381.

Jaeger, T. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59:434–446.

Johnson, K., Ladefoged, P., and Lindau, M. (1993). Individual differences in vowel production. *Journal of Acoustical Society of America*, 94:701–714.

Karlsson, F. (2004). Development of complex syllable onsets: Evidence from durational measurements. *Fonetik 2004, The XVIIth Swedish Phonetics Conference*, pages 100–103.

Karlsson, F. (2006). *The acquisition of contrast: A longitudinal investigation of initial s+plosive cluster development in Swedish children*. PhD thesis, Umea University, Umea.

Kent, R. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of Speech and Hearing Research*, 19:421–447.

Kent, R. (2000). Research on speech motor control and its disorders: A review and prospective. *Journal of Communication Disorders*, 33:391–428.

Kent, R. and Forner, L. (1980). Speech segment durations in sentence recitations by children and adults. *Journal of Phonetics*, 8:157–168.

Klatt, D. (1974). The duration of [s] in English words. *Journal of Speech, Language and Hearing Research*, 17:51–63.

Klatt, D. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59:1208–1221.

Kuehn, D. and Moll, K. (1976). A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics*, 4:303–320.

Lee, S., Potamianos, A., and Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America*, 105:1455–1468.

Lei, C., Fisher, S., Hurst, J., Vargha-Khadem, F., and Monaco, A. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, 413:519–523.

Levelt, W. (1989). *Speaking: From intention to articulation*. The MIT Press, Cambridge.

Lewis, B., Freebairn, L., Hansen, A., Taylor, H., Iyengar, S., and Shriberg, L. (2004). Family pedigrees of children with suspected childhood apraxia of speech. *Journal of Communication Disorders*, 37:157–175.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W.J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp.403–439). Kluwer, Dordrecht.

Maassen, B. (2002). Issues contrasting adult acquired versus developmental apraxia of speech. *Seminars in Speech and Language*, 23:257–266.

Maassen, B., Nijland, L., and Van Der Meulen, S. (2001). Coarticulation within and between syllables by children with developmental apraxia of speech. *Clinical Linguistics & Phonetics*, 15:145–150.

Marquardt, T., Jacks, A., and Davis, B. (2004). Token-to-token variability in developmental apraxia of speech: Three longitudinal case studies. *Clinical Linguistics & Phonetics*, 18:127–144.

Marquardt, T., Sussman, H., Snow, T., and Jacks, A. (2002). The integrity of the syllable in developmental apraxia of speech. *Journal of Communication Disorders*, 35:31–49.

Martin, K. and Ramnarine, K. (2003). Physics. In P.R. Hoskins & A. Thrush & K. Martin & T.A. Whittingham (Eds.), *Diagnostic ultrasound: Physics and equipment* (pp.7–22). Greenwich Medical Media, London.

McCabe, P., Rosenthal, J., and McLeod, S. (1998). Features of developmental dyspraxia in the general speech impaired population? *Clinical Linguistics & Phonetics*, 12:105–126.

Meyer, A. (1994). Timing in sentence production. *Journal of Memory and Language*, 33:471–492.

Morley, M., Court, D., and Miller, H. (1954). Developmental dysarthria. *British Medical Journal*, 1:8–10.

Munson, B., Bjorum, E., and Windsor, J. (2003). Acoustic and perceptual correlates of stess in nonwords produced by children with suspected developmental apraxia of speech and children with phonological disorder. *Journal of Speech, Language and Hearing Research*, 46:189–202.

Nakagawa, S. (2004). A farewell to Bonferroni: The problems of low statistical power and publication bias. *Behavioral Ecology*, 15:1044–1045.

Nijland, L. (2009). Speech perception in children with speech output disorders. *Clinical Linguistics & Phonetics*, 23:222–239.

Nijland, L., Maassen, B., Hulstijn, W., and Peters, H. (2004). Speech motor coordination in Dutch-speaking children with DAS studied with EMMA. *Journal of Multilingual Communication Disorders*, 2:50–60.

Nijland, L., Maassen, B., and Van Der Meulen, S. (2003a). Evidence of motor programming deficits in children diagnosed with DAS. *Journal of Speech, Language and Hearing Research*, 46:437–450.

Nijland, L., Maassen, B., van der Meulen, S., Gabreels, F., Kraaimaat, F., and Schreuder, R. (2002). Coarticulation patterns in children with developmental apraxia of speech. *Clinical Linguistics & Phonetics*, 16:461–483.

Nijland, L., Maassen, B., Van Der Meulen, S., Gabreels, F., Kraaimaat, F., and Schreuder, R. (2003b). Planning of syllables in children with development apraxia of speech. *Clinical Linguistics & Phonetics*, 17:1–24.

Nittrouer, S. (1993). The emergence of mature patterns is not uniform: Evidence from an acoustic study. *Journal of Speech and Hearing Research*, 36:959–972.

O'Shaughnessy, D. (1974). Consonant durations in clusters. *IEEE Transactions on Acoustics Speech and Signal Processing*, 22:282–295.

Ostry, D., Feltham, R., and Munhall, K. (1984). Characteristics of speech motor development in children. *Developmental Psychology*, 20:859–871.

Peter, B. and Stoel-Gammon, C. (2005). Timing errors in two children with suspected childhood apraxia of speech (sCAP) during speech and music-related tasks. *Clinical Linguistics & Phonetics*, 19:67–87.

Peter, B. and Stoel-Gammon, C. (2008). Central timing deficits in subtypes of primary speech disorders. *Clinical Linguistics & Phonetics*, 22:171–198.

Pollock, K. and Hall, P. (1991). An analysis of the vowel misarticulations of five children with developmental apraxia of speech. *Clinical Linguistics & Phonetics*, 5:207–224.

R Development Core Team (2009). R: A language and environment for statistical computing. Retrieved 15 June 2009 from http://www.R-project.org.

Robbins, J. and Klee, T. (1987). Clinical assessment of oropharyngeal motor development in young children. *Journal of Speech and Hearing Disorders*, 52:271–277.

Saville, D. J. (1990). Multiple comparison procedures: The practical solution. *The American Statistician*, 44:174–180.

Shawker, T. and Sonies, B. (1985). Ultrasound biofeedback for speech training: Instrumentation and preliminary results. *Investigative Radiology*, 20:90–93.

Shriberg, L. and Aram, D. (1997a). Developmental apraxia of speech: II. Toward a diagnostic marker. *Journal of Speech, Language and Hearing Research*, 40:286–313.

Shriberg, L. and Aram, D. (1997b). Developmental apraxia of speech: III. A subtype marked by inappropriate stress. *Journal of Speech, Language and Hearing Research*, 40:313–337.

Shriberg, L., Ballard, K., Tomblin, B., Duffy, J., Odell, K., and Williams, C. (2006). Speech, prosody, and voice characteristics of a mother and daughter with 7;13 translocation affecting FOXP2. *Journal of Speech, Language and Hearing Research*, 49:500–525.

Shriberg, L., Campbell, T., Karlsson, H., Brown, R., Mcsweeny, J., and Nadler, C. (2003a). A diagnostic marker for childhood apraxia of speech: The lexical stress ratio. *Clinical Linguistics & Phonetics*, 17:549–574.

Shriberg, L., Green, J., Campbell, T., Mcsweeny, J., and Scheer, A. (2003b). A diagnostic marker for childhood apraxia of speech: The coefficient of variation ratio. *Clinical Linguistics & Phonetics*, 17:575–595.

Smit, A., Hand, L., Freilinger, J., Bernthal, J., and Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders*, 55:779–798.

Smith, B. (1978). Temporal aspects of English speech production: A developmental perspective. *Journal of Phonetics*, 6:37–67.

Smith, B. and Kenney, M. (1998). An assessment of several acoustic parameters in children's speech production development: Longitudinal data. *Journal of Phonetics*, 26:95–108.

Stone, M. (1990). A three-dimensional model of tongue movement based on ultrasound and x-ray microbeam data. *Journal of the Acoustical Society of America*, 87:2207–2217.

Stone, M. (2005). A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics & Phonetics*, 19:455–501.

Stone, M., Epstein, M., and Iskarous, K. (2004). Functional segments in tongue movement. *Clinical Linguistics & Phonetics*, 18:507–521.

Stone, M. and Lundberg, A. (1996). Three-dimensional tongue surface shapes of English consonants and vowels. *Journal of the Acoustical Society of America*, 99:3728–3737.

Sussman, H., Marquardt, T., Doyle, J., and Knapp, H. (2002). Phonemic integrity and contrastiveness in developmental apraxia of speech. In F. Windsor & L. Kelly & N. Hewlett (Eds.), *Investigations in clinical phonetics and linguistics* (pp.311–326). Lawrence Erlbaum Associates, Mahwah.

Templin, M. (1957). *Certain language skills in children*. University of Minessota Press, Minneapolis.

Thoonen, G., Maassen, B., Gabreels, F., and Schreuder, R. (1999). Validity of maximum performance tasks to diagnose motor speech disorders in children. *Clinical Linguistics & Phonetics*, 13:1–23.

Thoonen, G., Maassen, B., Gabreels, F., Schreuder, R., and deSwart, B. (1997). Towards a standardised assessment procedure for developmental apraxia of speech. *European Journal of Disorders of Communication*, 32:37–60.

Umeda, N. (1977). Consonant duration in American English. *Journal of the Acoustical Society of America*, 61:846–858.

Vargha-Khadem, F., Gadian, D., Copp, A., and Mishkin, M. (2005). FOXP2 and the neuroanatomy of speech and language. *Neuroscience*, 6:131–138.

Vazquez Alvarez, Y. and Hewlett, N. (2007). The "trough effect": An ultrasound study. *Phonetica*, 65:105–121.

Whittingham, T. (2003). Transducers and beam-forming. In P.R. Hoskins & A. Thrush & K. Martin & T.A. Whittingham (Eds.), *Diagnostic ultrasound: Physics and equipment* (pp.23–48). Greenwich Medical Media, London.

Wrench, A. and Scobbie, J. (2006). Spatio-temporal inaccuracies of video-based ultrasound images of the tongue. In *Proceedings of the 7th International Seminar on Speech Production, Brazil (ISSP7). CD-ROM*.

Zharkova, N. and Hewlett, N. (2009). Measuring lingual coarticulation from midsagittal tongue contours: Description and example calculations using English /t/ and /a/. *Journal of Phonetics*, 37:248–256.